Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
HERSCHEL W. LEIBOWITZ, *International Business Machines Corporation*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

A Factorial Analysis of Verbal Learning Tasks: P. A. GAMES 1
A Study of the Effects of Verbalization on Problem Solving: R. M. GAGNÉ AND E. C. SMITH, JR. 12
Countertraining of a Simple Skill with Immediate and 1-Week Delays of Informative Feedback: F. J. RYAN AND E. A. BILODEAU 19
A Test of the Response Probability Theory of Perceptual Defense: M. J. GOLDSTEIN 23
Predicting Instrumental Performance from the Independent Rates of Contingent Responses in a Choice Situation: A. J. BROWNSTEIN 29
Magnitude of Reward and Probability Learning: Y. BRACKBILL, M. S. KAPPY, AND R. H. STARR 32
The Influence of Strength of Drive on Functional Fixedness and Perceptual Recognition: S. GLUCKSBERG 36
Preference Probability between Gambles as a Step Function of Event Probability: R. D. LUCE AND E. F. SHIPLEY 42
Extinction Effects Following Nondifferential Reinforcement of an Irrelevant Stimulus: S. E. SPERLING 50
The Role of Drive (Time Stress) in Complex Learning: An Emphasis on Prelearning Phenomena: R. E. CLARK 57
The Effectiveness of Concepts at Various Levels of Awareness: K. G. DAVIS AND H. F. HESS 62
Size-Constancy Judgments and Perceptual Compromise: V. R. CARLSON 68
The Operation of Set in a Visual Search Task: H. FOSTER 74
Anticipatory Timing of Continuous and Discrete Responses: J. A. ADAMS AND L. R. CREAMER 84
Shape-Constancy: Dependence upon Stimulus Familiarity: C. R. BORRESEN AND W. H. LICHTE 91
Latent Extinction as a Function of Number and Duration of Pre-Extinction Exposures: J. A. DYAL 98

American Psychological Association

Vol. 63, No. 1 January 1962

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $10.00, or $20.00 annually. Single copies are $2.00. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, 53, 1–2).

Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.* When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1961 the publication cost is $30 per printed page.

All **regular** articles, **supplementary reports,** and **replication reports** should be submitted to the Editor-elect: David A. Grant: Department of Psychology, University of Wisconsin; Madison 6, Wisconsin. **Early publication** articles should be sent to the Editor: Arthur W. Melton; Department of Psychology, University of Michigan; Ann Arbor, Michigan until *July 1, 1962.* Thereafter they should be sent to the Editor-elect. Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts.* Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR
Acting Production Manager

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1962 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 63, NO. 1 JANUARY 1962

A FACTORIAL ANALYSIS OF VERBAL LEARNING TASKS[1]

PAUL A. GAMES

Ohio University

The methods of factor analysis have been applied recently to measures typically associated with the experimental laboratory in order to identify sources of common variance on such tasks and to trace the changes in factorial structure associated with practice. Fleishman (1957, 1960) and Fleishman and Hempel (1954, 1955) have demonstrated, in the psychomotor area, that the factorial composition of a task may substantially change with practice. This study is an attempt to determine the extent of common variance among some verbal learning tasks and to identify the common factors that the tasks are measuring. More specifically, the type of task (paired-associate or serial list), method of presentation and response, and order of presentation were varied to see if these were related to the abilities measured. Also investigated was the extent to which the contributions of the ability factors change with early practice on verbal learning tasks.

The identification of sources of variance is greatly facilitated if reference tests are used to define factors whose interpretation, from the accumulated results of previous studies, is relatively clear. One factor clearly related to verbal learning is "Rote Memory." This factor is defined (Christal, 1958) by ". . . tests requiring the formation of associations . . ." (p. 1) and has been reported in at least 17 factor analysis studies (French, 1951, p. 219). The factor appears to transcend the "content" of the stimulus-response pairs "to be associated" since tasks involving numbers, forms, words, letters, pictures, colors, and names have all appeared on the factor. The factor has been defined by both recall and recognition response measures, and with both paired associates and single lists (of items) as the material to be learned. However, the rote memory group tests have typically utilized a present-recall task (present pair 1 . . . present 10; recall response to pair 1 . . . recall 10) rather than the anticipation method commonly

[1] Based on work submitted in partial fulfillment of the requirements for the PhD degree at the State University of Iowa. The author is indebted to Harold P. Bechtoldt for advice and assistance during the course of the investigation.

used in experimental studies of verbal learning.

The serial anticipation task of verbal learning clearly involves the formation of associations. However, in terms of formal task similarity, such tasks have no counterpart in tests of Rote Memory, but are similar (on the early trials) to Span Memory tests. Span tests require the accurate reproduction of temporal sequences of stimuli. A Span Memory factor has been rather clearly identified in at least two studies (Brener, 1940; Woodrow, 1939) and suggested by the results of others (French, 1951, p. 240); however, sufficient measures to define both the Rote Memory and Span Memory factors had never been included in the same analysis, until Kelley (1954) rather clearly demonstrated the linear independence of the

TABLE 1

THE TEST BATTERY

Test No.	Test Name and Mode of Presentation	Stimulus Exposure Time (Sec/ Stimulus)	Response Time (Sec/ Stimulus)	Time between Trials (Sec.)	Total Time (Approx. Minutes)
	Day 1: Group Tests				
0′	Peripheral Span (flash cards)	.5	a	NA	3
1	Letter Span (flash cards)	2	a	NA	10
2	Symbol-Numbers (film)	4	6	8	10
3	Repeated Span (film)	1	20b	2	6
4	Word Memory (auditory)	2	4	NA	7
5	Random Pairs—Present-Recall (film)	2	2	4	7
	Free Recall				1
	Day 2: Individual Tests				
A	Warm-up Task (Missouri drum)	2	2	0	4
6	Serial Task-Anticipation (Missouri drum)	2	2	6	6
	Free Recall				1
7	Random Pairs—Present-Recall (Missouri drum)	2	2	6	6
8	Random Pairs—Anticipation (Hull drum)	2	2	4	6
	Free Recall				1
9	Repeated Span (Missouri drum)	2	20b	0	5
10	Constant Pairs—Anticipation (Hull drum)	2	2	4	4
	Free Recall				1
11	Constant Pair—Present-Free Recall (Missouri drum)	2	20b	0	6
	Day 3: Group Tests				
0″	Peripheral Span (flash cards)	.5	a	NA	3
12	S-2 Letter Span (film)	1.5	a	NA	8
13	R-3 First Names (flash cards)	4	6	NA	10
14	Constant Pairs: Present-Recall (film)	2	2	4	7
	Free Recall				1
15	Number Span (auditory)	1	a	NA	6
16	Random Serial Pairs: Present-Recall (film)	2	2	4	7
	Free Recall				1

Note.—NA = not applicable; a = see text; b = total response time allowed, no separate stimuli presented.

two factors. The early trials on a serial task thus might show a high loading on the Span factor, while the later trials shift toward the Rote Memory factor as new associations are formed.

In addition to the present-recall vs. anticipation method differentiation, verbal learning tasks vary in terms of the constancy of the sequence of presentation. Serial tasks, of necessity, have a constant order over all trials; whereas paired-associate tasks may or may not have a constant order. Another differentiation between the serial and paired-associate tasks is that in the former, each item serves as both stimulus and response. Any or all of these differentiations may affect the loadings on the Rote Memory or Span Memory factors, or may be associated with systematic individual differences, and hence may define separate factors. To investigate these possibilities and the changes in factor loadings associated with practice, the following specific hypotheses were formulated, and the battery of 17 tests listed in Table 1 was developed to investigate these hypotheses.

Hypotheses Tested

Prerequisite hypothesis.—Hypothesis 1 was that four of the tests used by Kelley (1954) and two tests used by other investigators (French, 1951) define two linearly independent factors—Rote Memory (Tests 2, 4, and 13) and Span Memory (Tests 1, 12, and 15). These measures are designated the "reference tests." All other measures are referred to as the "experimental tests." This hypothesis was labeled prerequisite in the sense that the following hypotheses, as stated, could be tested only if this hypothesis were retained. This hypothesis was tested by Part I of the analysis.

Major hypothesis.—Hypothesis 2 was that there are positive correlations between the reference and experimental tests and these correlations are accounted for by the two reference factors, Rote Memory and Span Memory. This hypothesis was tested, using the correlations between the reference and experimental tasks, (*a*) by the magnitude of the observed correlations, and (*b*) by the insignificance of residuals after the effects of the two reference factors have been eliminated.

Supplemental hypotheses.—One or more common factors *restricted to the experimental tasks* may or may not appear even if the major hypothesis is retained. Such factors could be compatible with the major hypothesis. The following possible factors were suggested by the preceding analysis. Tests that would define such factors were represented in the design of the study.

Hypothesis 3a was that a group factor is defined by the anticipation method tasks, (Tests 6, 8, and 10). Hypothesis 3b was that a group factor is defined by the constant order tasks (Tests 3, 6, 9, 10, 11, and 14). Hypothesis 3c was that a group factor is defined by the repeated Span and Serial tasks (Tests 3, 6, and 9). These hypotheses were tested by Part II of the analysis.

Special hypotheses.—Hypothesis 4a was that scores on successive trials on the serial task systematically shift, with practice, from the Span factor to the Rote Memory factor. Hypothesis 4b was that free recall scores on the experimental tasks have higher loadings on the Span factor than the regular anticipation or present-recall scores. These hypotheses were tested by Part III of the analysis. The specific manner in which these hypotheses were tested is given in the procedure below.

METHOD

Subjects

The sample consisted of 100 volunteers from the elementary psychology course at the State University of Iowa. To prevent conflicts in scheduling with another verbal learning experiment and to insure that no *Ss* had prior experience on verbal learning tasks, *Ss* were selected from those having scores be-

tween Percentiles 20 and 80 on the alow Picture Interpretation Test. No significant differences were found between these upper and lower groups on the verbal paired-associate task in this other experiment. Thus, there is no indication that this selection affected the factorial results.

Procedure

The test battery shown in Table 1 was administered in three sessions held within 1 week. The groups in Sessions 1 and 3 ranged in size from 6 to 20 *Ss*. Session 2 was an individual session. All tests were administered by *E*.

Group tests of auditory presentation were administered by means of a tape recorder placed in the front of the room. Film strip tests were given by use of a Soundview automatic projector and two Hunter-Brown timers, and were projected on a screen at the front of the room. These timers were left on throughout all parts of the group session (except when the auditory tests were given) to provide for the timing of the flash card tests. Written responses were required on all group tests. Oral responses were obtained on the individual tests administered by means of a Hull-type and a Missouri-type memory drum.

Test Battery

Table 1 lists the tests, the order and mode of presentation of the tests, the stimulus exposure time, the time allowed for a response, and the time between trials when applicable.

Warm-up tasks.—The warm-up task for both group sessions was a peripheral Span test in which a sequence of numbers was presented on a single flash card. Seven sequences of random numbers ranging from 4 to 10 digits in length were presented in order of increasing length. The number of completely correct sequences in both tasks (0′ and 0″) constituted Variable 0.

The warm-up task for the individual session, designed to familiarize *Ss* with the anticipation method, consisted of four trials on a list of five pairs of typewriter symbols and numbers presented in a constant order. If the subject was not responding according to the instructions by the end of Trial 4, he was given two more trials.

Span Memory reference tests.—The three Span Memory reference tests allowed 10 sec. for response at the end of sequences of five items or less, 15 sec. for sequences of six to eight items, and 20 sec. for sequences over eight items. The three reference tests are slight modifications of Kelley's (1954) Span tests.

Test 1: Letter Span. After seeing a sequence of letters one at a time, *S* must write the sequence. Two test series of six items each were presented, the items ranging from 5 to 10 letters in length. Within each test series the six items were arranged in increasing order of length. Two sequences of 4 letters in length preceded the first test sequence. The score was the number of sequences completely correct.

Test 12: Letter Span. Same as Test 1 except presentation of letter sequences was done by film strip rather than flash cards.

Test 15: Number-Span. Same as Test 1 except presentation of number sequences was by a tape recorder.

Rote Memory reference tests.—The score on each of these tests is the number of correct responses.

Test 2: Symbol-Numbers. Ten pairs of simple figures were associated with two-digit numbers. Each of the three trials consisted of a presentation of all pairs and a test of each stimulus alone. Preceding the test, *Ss* were given a practice task of four pairs to familiarize them with the procedure.

Test 4: Word Memory. This is Kelley's (1954) "Memory for Words I." Pairs of unrelated one-syllable nouns were read aloud twice, then the first word of each pair was read. The 20-item test was given in two parts, administered consecutively with 10 pairs per part.

Test 13: First Names. This is a task of associating first and last names. The test consisted of a learning presentation of a list of 10 pairs of first and last names, and a test presentation of the last names only, followed by another presentation of both names and another test presentation. A second list of 10 names was then given with two learning and two test presentations as above.

Experimental tests.—To reduce the possibility of specific content factors that might prevent a test of the specified hypotheses, all experimental tests used only the 21 consonant letters as stimuli and responses. Letters were chosen because of high initial familiarity and response availability that should increase rate of learning and that might reduce any "learning to learn" effects. All serial and span tests consisted of a list of 10 consonants while paired-associate tasks involved 10 pairs, or 20 consonants. Estimates of initial associational values of pairs of consonants have been based upon data from an unpublished study by Bechtoldt (1956) in which 90 *Ss* gave two single-letter

responses to each consonant. Only one association that had received a total frequency over seven in the above study has been used as a correct association in the lists for this experiment. The *E* also attempted to avoid any associations that might have been learned in previous tasks in this study or acquired from experiences outside the laboratory; (e.g., PT as in PT boats, etc.). (Test 11 was added to the experimental battery last; no list was found that could meet the criterion listed, so several pairs of Test 5 were reversed and used.)

Since the correct associational strength has been kept low, but high associations of each stimulus with other incorrect letters remain, these tests may be conceived of as low association-high competition tasks in contrast to the reference tests that are best described as low association-low competition tasks.

Six trials were given on most experimental tasks with three scores obtained from each set of trials; the three scores were defined as the total number of correct responses on successive blocks of two trials, i.e., Trials 1 and 2, 3 and 4, and 5 and 6. These three block scores were designated the a, b, and c measures, respectively. Free recall tests, in which *S* was asked to write down, or recite as many S-R pairs as he could recall, were designated as the d measure when given. The number of trials and the usage of a final free recall test can be determined for each test by consulting Table 3. (Trials 5 and 6 of Test 10 were run, but dropped from the analysis because of an extreme ceiling effect.) All group experimental tests were given by film strip; all individual tests, by memory drum.

Repeated Span and Serial tests.[2]—Test 3: Repeated Span. Each of the four trials consisted of a ready signal followed by the separate exposure of each of 10 consonants with 20 sec. allowed at the end of the list for *Ss* to write the sequence.

Test 9: Repeated Span. Same as Test 3 with 10 blanks on the memory drum giving *S* 20 sec. to verbally recite the sequence. The last blank contained an asterisk as the ready signal for the coming trial.

Test 6: Serial Task. The one presentation trial was followed by six test trials using the anticipation method. An asterisk after the last letter indicated the start of a new trial and served as the cue for the anticipation of the first letter of the list.

Experimental Letter-Pair tasks.—Since most of these tests differ in only two ways (constant order of pairs and test stimuli over trials vs. random order; and method of presenting and testing), they may be described by title alone. The anticipation method tasks all involved one presentation trial prior to the prescribed number of test trials; a present-recall trial consisted of a presentation series followed by a recall series with each stimulus presented alone. The present-free recall task differs from the above only in that the recall series was an interval in which *S* attempted to write down as many pairs as he could.

Test 5: Random Pairs; Present-Recall (film strip)

Test 7: Random Pairs; Present-Recall (Missouri memory drum)

Test 8: Random Pairs; Anticipation (Hull memory drum)

Test 10: Constant Pairs; Anticipation (Hull memory drum)

Test 11: Constant Pairs; Present Free Recall (film strip)

Test 14: Constant Pairs; Present Recall (film strip)

Test 16: Random Serial Pairs; Present Recall (film strip). These pairs were constructed by taking adjacent letters on a serial list of 10 letters as pairs, plus the tenth letter and the first letter as a pair.

Analysis of Data

Product-moment correlation coefficients were computed for all pairs of measures using IBM punched card procedures. The analysis of the correlations involved three parts:

1. The factoring of the reference tests (1, 2, 4, 12, 13, 15).

2. The projecting of the second-block scores of the experimental tests (3b, 5b, 6b, 7b, 8b, 9b, 10b, 11b, 14b, 16b) upon the factors defined in Part 1, and the factoring of the residual matrix of these measures.

3. The projection of the free association trials and first and third blocks of the experimental tests onto the factors defined in Parts 1 and 2 (3a, 5a, 5c, 6a, 6c, 6d, 7a, 7c, 8a, 8c, 8d, 9a, 10a, 11a, 11c, 14a, 14d, 16a, 16d, 0). Table 2 designates submatrices of the complete correlation matrix according to this division.

[2] On Tests 3, 9, and the free recall measure of Test 6, the number of correct responses was defined as the number of consonants reproduced on the correct spaces of the answer sheet; however, if a group of three or more consonants in correct order was incorrectly located, the group was scored as the number of correct letters minus the number of spaces the group was displaced.

TABLE 2

THE COMPLETE CORRELATION SUPERMATRIX

Measures	Reference Tests	Experimental Tests	Supplementary Measures
Reference tests	1	2	4
Experimental tests (b measures only)	2′	3	5
Supplementary measures of experimental tests (a, c, and d measures)	4′	5′	6

The projection procedure in Parts 2 and 3 was the following variation of the multiple-group factoring technique originated by Bechtoldt (1956):

Let R_1 = the matrix of intercorrelations of the "reference" variables defining the factor space; the set of "reference" variables for Part 3 will include the reference tests of Part 1 and the variables of Part 2.

Let R_2 = the matrix of correlations of the "variables to be projected" with the appropriate set of "reference" variables.

Let E = the weighing matrix with entries of $+1$ or -1 as determined by the centroid factoring of R_1;

Define and compute:

$T = ER_1E'$, a symmetric matrix

C, a triangular matrix, where $CC' = T$ and C is defined by the diagonal factoring steps of the multiple group method of factoring.

$(C')^{-1}$ = the inverse of the transpose of C,

Then $F_2 = R_2E'(C')^{-1}$.

The entries of F_2 are the factor loadings of the "projected" variables on the orthogonal centroid factors defined by the "reference" variables.

1. The centroid factoring of the Part 1 or reference variables (Matrix 1) yielded two factors (I and II) significant by a criterion based on five approximation procedures.[3] The third factor extracted was rejected by four of five tests, therefore only two factors were retained. One cycle of multiple group factoring was necessary to stabilize each communality to within $\pm.03$ of its value the previous cycle.

2. The loadings of the experimental tests on Factors I and II were determined by the projection method using the data of Matrix 2. The matrix of two factor residuals of the intercorrelations of the experimental tests, Matrix 3, was computed using the loadings on Factors I and II, by the usual R-FF' procedure. Centroid factoring of the residuals yielded three additional factors (III, IV, and V.) Factor V was significant on three of the five approximate tests and therefore was retained. Three cycles of the multiple group method were necessary to stabilize the communalities.

3. The loadings of the remaining measures on Factors I and II were determined by the data of Matrix 4, and the loadings on Factors III, IV, and V were determined by the correlations of Matrix 5. In the latter matrix some of the correlations are intertrial correlations on the same task and therefore include "specific factor" variance (e.g., the correlation of 3a with 3b). Since this specific variance is to be excluded from the analysis it is necessary to derive stable estimates of the intertrial correlations based on the common factors of the battery. This estimate was secured by three successive cycles of the multiple group method.

To test the hypotheses previously specified, Factors I and II were rotated only with respect to each other, i.e., they were kept orthogonal to the set of Factors III, IV, and V defined by the experimental tests alone. Similarly Factors III, IV, and V were rotated with respect to each other only. Within these limitations the factors were graphically rotated to an oblique positive simple structure in accordance with Thurstone's (1947) criterion. Loadings on all tests were used in determining the location of the factors.

RESULTS AND INTERPRETATIONS

The factor loadings of all tests on the five rotated factors, the portion of the variance accounted for by the two reference factors, and the mean and variance of the raw scores for each measure are given in Table 3.[4]

[3] The criterion of significance adopted consisted of the application of five approximation tests of significance—first those of Mosier, Guilford and Lacy, and McNemar; if all yielded the same result, no further tests were run. If one of the tests differed from the other two, two additional tests were run, Tucker's and Burt's (Thompson, 1951, pp. 121–123).

[4] The complete orthogonal F matrix, the transformation matrix, and the product-moment correlations, and fifth factor residuals for all variables have been deposited with the American Documentation Institute. Order Document No. 6905 from ADI Auxiliary

TABLE 3

ROTATED FACTOR LOADINGS, COMMUNALITIES ON THE REFERENCE FACTORS, MEANS, AND VARIANCES OF THE VARIABLES

Test No.	Test Name and Mode of Presentation	Factors					Reference Factor Communality	Mean	σ^2
		SM(I)	RM(II)	III	IV	V			
2	Letter Span (cards)	72	04	−03	03	09	60	4.40	4.18
12	Letter Span (film)	76	12	−09	−04	−05	72	4.28	4.61
15	Number Span (auditory)	60	04	12	04	−04	43	4.43	5.51
0	Peripheral Span (cards)	63	−03	−02	08	22	42	8.05	3.69
2	Symbol-Numbers (film)	−06	69	−11	−03	10	51	13.66	36.58
4	Word Memory (auditory)	−02	72	05	04	−04	57	12.07	23.17
13	First Names (cards)	03	70	05	00	−07	55	24.48	65.51
3a	Repeated Span (film)	60	−10	−12	−15	28	36	9.92	11.09
3b	Same	42	27	−04	04	59	35	15.04	14.90
9a	Repeated Span (drum)	48	−10	12	31	25	19	9.36	13.57
9b	Same	30	29	05	19	30	26	14.07	23.27
5a	Random Pairs: Present-Recall (film)	28	42	30	20	−02	38	5.55	12.75
5b	Same	11	59	30	−10	23	45	11.51	29.01
5c	Same	22	56	36	−05	19	49	14.22	28.19
6a	Serial Task: Anticipation (drum)	−05	58	−04	35	07	36	5.23	12.68
6b	Same	01	48	01	46	−02	26	9.66	20.72
6c	Same	−06	54	08	56	06	31	12.38	22.54
6d	Same	−04	60	−04	43	−01	38	6.88	7.41
7a	Random Pairs: Present-Recall (drum)	20	45	33	13	−02	33	5.38	12.20
7b	Same	26	58	39	01	−02	54	10.40	21.78
7c	Same	12	64	42	06	12	52	12.79	20.91
8a	Random Pairs: Anticipation (drum)	24	42	56	16	−01	33	3.44	9.39
8b	Same	19	40	61	18	02	27	7.53	21.55
8c	Same	21	47	57	25	05	35	10.81	28.77
8d	Same	16	44	54	24	04	29	5.26	5.35
10a	Constant Pairs: Anticipation (drum)	32	34	45	22	04	32	5.32	11.24
10b	Same	21	47	45	30	16	26	10.92	22.13
10d	Same	21	52	26	23	18	42	5.32	4.86
11a	Constant Pairs: Present-Free Recall (drum)	14	38	16	04	30	22	7.03	6.31
11b	Same	04	53	22	16	35	33	11.25	9.81
11c	Same	11	41	28	07	23	22	13.60	12.70
14a	Constant Pairs: Present Recall (film)	24	38	51	08	08	30	9.98	14.00
14b	Same	05	44	40	08	19	23	16.81	13.59
14d	Same	14	29	37	15	09	15	8.03	3.79
16a	Random Serial Pairs: Present-Recall (film)	08	32	34	05	12	14	5.13	6.33
16b	Same	08	51	54	−05	−07	33	9.70	20.45
16c	Same	08	50	44	−05	04	30	11.23	24.72
16d	Same	09	45	35	13	23	26	5.64	5.31

Note.—Decimals omitted from factor loadings and communalities.

Prerequisite hypothesis.—Hypothesis 1 was that the reference tests define two linearly independent factors, Rote Memory (RM) and Span Memory (SM). This hypothesis was retained. The Span reference tests displayed high loadings on the first factor and the Rote Memory reference variables were highest on the second factor. The correlation between the Rote Memory and Span Memory factors was .32.

High correlations were found between the Peripheral Span task and the temporal span tasks, so that the ensuing projection of Task O on the Span factor was as high as that of one reference variable. Although the mean score of the Peripheral Span task was considerably higher than that of the temporal span tests, the factorial composition was essentially the same.

Major hypothesis.—Hypothesis 2 was that there are positive correlations between the reference and experimental tests, and these correlations are accounted for by the two reference factors. This hypothesis was retained. All but one of the 186 correlations between the experimental

Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

and reference tasks were positive and 108 of these correlations were above .254, the r corresponding to the 1% level of significance. The adequacy of the two reference factors was demonstrated by the lack of significance of the two factor residuals of Matrix 2. All experimental variables showed loadings of .29 or higher on one or both of the reference factors.

The Letter Pair tests all displayed their maximal loadings on the Rote Memory factor, as expected, and had very low loadings on the Span factor. However, the median loading of the Rote Memory reference tests was .70, while the median loading of all Letter Pair measures was .47. This decrease may have been partly due to the low association-low competition nature of the reference tests and the contrasting low association-high competition character of the Letter Pair tasks.

The median loading of the Letter Pair tests on the Span factor was .20. While the factor loadings of the Letter Pair tests were relatively low, the values were certainly not centered around the expected zero value. One possible explanation of this result may lie in the composition of the reference tests. The Span reference tests consisted of two Letter Span tests presented visually and only one Number Span test with auditory presentation. Hence, any "visual letter discrimination or general letter content" source of test variance was common to the two Span reference tests, and was, therefore, included in that factor, while other "content" characteristics were not. The fact that the Letter Span tests had communalities of .70 and .62 in the two-factor reference space while the Number Span and Peripheral Number Span tests had communalities of .43 and .42, respectively, support the interpretation of a slight bias of the Span factor toward letter content. Such a content bias would very easily account for the consistent, but low positive loadings of the Letter Pair tests.

Supplemental hypotheses.—Three common factors (III, IV, and V) restricted to the experimental tasks were extracted and left orthogonal to the reference factors. None of the reference tests had loadings above .12 on any of these factors. It will be recalled that the significance of Factor V (three of five approximate tests yielded significant values) and its retention was a somewhat arbitrary decision. Therefore two rotations were made, one with Factors III, IV, and V the other with only Factors III and IV. The loadings on Factor III were about the same in both solutions, the largest difference being .08 and the mean of the absolute differences being only .018. The loadings on Factor III for the five-factor solution are found in Table 3.

All Letter Pair tasks had loadings above .20 on Factor III. The only zero loadings were for the reference tests, the serial task and the repeated Span tasks. The factor was therefore specific to letter pair tasks, and was designated as a "Letter Pairs Content Factor." It is suggested, on the basis of *Ss*' verbal reports, that the factor represents a fluency in making associations between letter pairs. This interpretation is supported by the apparent reduction in loadings accompanying the use of the free recall procedure, since if the stimulus is not presented to elicit the created association, the variance associated with this factor should be reduced. No simple formulation, however, seems to account for the other differential loadings on this

factor, and any complex formulation on the basis of these limited data would be of questionable value.

Factor IV failed to account for much of the variance of the test battery. It also failed to demonstrate the stability of Factor III in the two rotational solutions; its loadings changed appreciably. The five-factor solution suggested an anticipation method factor as in Hypothesis 3a, since 9 of the 10 loadings above .20 were on anticipation method tasks (6, 8, and 10), the only exception being a loading of .31 on the first block of the individual Repeated Span task (9a). The four-factor solution, however, is not as readily interpretated, the loadings above .20 on Factor IV were: Test 3b, .47; Test 9a, .38; Test 9b, .36; Test 6a, .29; Test 6c, .40; Test 6d, .27; Test 10d, .28; Test 11a, .25; Test 11b, .37; Test 11c, .21; Test 16d, .25; Test 0, .22.

The presence of loadings on Tests 3, 9, and 6 suggests that a group factor is defined by the Repeated Span and Serial Tasks (Hypothesis 3c), but the other loadings suggest a greater generality. Hypothesis 3b, that a factor is defined by the constant-order tasks, is supported by the fact that Tests 3, 6, 9, and 11 all have a constant presentation order. However if this is the crucial differentiating factor, where are the a, b, and c measures of Tests 10 and 14, which also involve a constant order? Thus no well supported statement may be made about Hypotheses 3a, 3b, and 3c, except that at least one of the three is false. Which, if any, is true will have to be investigated in future studies.

Factor V, in the five-factor solution also accounted for relatively little test variance, and had only one measure with a large loading; the second measure of the group Repeated Span test (3b) had a loading of .59. The other repeated span measures and the Constant Pairs-Free Recall measures all had loadings between .23 and .35, suggesting that the free recall mode of response might define the factor. However, the free recall trials of the other tests failed to show substantial loadings on Factor V; these loadings were −.01 (6d), .04 (8d), .09 (14d), and .23 (16d). Thus no interpretation seemed reasonably supported.

Special hypotheses.—Hypothesis 4a stated that scores on the serial learning task shift from the Span factor to the Rote Memory factor with practice. This hypothesis was rejected; the Serial task displayed high Rote Memory loadings on all blocks of trials and approximately zero loadings on the Span factor on all blocks. Hypothesis 4b, that free recall scores on the experimental tasks have higher loadings on the Span factor than the regular anticipation or present recall scores was also rejected.

The Repeated Span tests (3 and 9), however, did exhibit a consistent change in factor loadings with practice. The loadings on these tests shifted from the Span factor toward the Rote Memory factor as the amount of practice was increased. These measures clearly changed, with practice, from relatively pure span measures to a composite of the two factors. This change of factorial structure with practice was the only large consistent change observed with respect to the reference factors. There was, however, a suggestion of a small change in the Letter Pair tests. As practice increased from the trials of Block 1 to those of Block 3, the mean loadings decreased on the Span factor (from .21 to .15) and increased

on the Rote Memory factor (from .39 to .52); at the same time, a small increase in the two factor communalities occurred (from .29 to .38). The individual tests, however, did not all demonstrate this trend, so that the most parsimonious explanation is probably to attribute the Letter Pair tests shifts in factor loadings to chance, unless this finding is replicated in another study.

Discussion

Christal (1958) offers two alternative hypotheses about the Rote Memory factor. "The first is that associative memory is a rather limited and artificial factor which reflects the extent to which examinees resort to use of 'memory crutches' in a formal testing situation" (p. 22). According to this interpretation, there should be a systematic reduction in Rote Memory loadings for free recall measures, since the absence of the initial stimulus portion of the "memory crutch" should greatly reduce the usefulness of this device. No such reduction was found.

"An alternative hypotheses . . . is that it [Rote Memory] represents ability to reproduce stimuli after limited exposure. This ability may have little relationship to the ability to reproduce the same materials after they have been well integrated through many exposures under various conditions and various contexts" (p. 22). The data from this study are compatable with this hypothesis as long as six trials on a verbal learning task are considered as a limited exposure. This hypothesis clearly implies that as learning progresses, loadings will shift away from the Rote Memory factor, but when this would occur is unspecified. No decrease in Rote Memory loadings from Trials 1 to 6 was observed in this experiment.

In either case, Rote Memory would seem to be the major individual difference learning parameter needed in theories of behavior in verbal learning situations. Approximately one-fifth to one-fourth of the total variance of the Letter Pair tasks and nearly one-third of the total variance of the Serial Task can be attributed to the Rote Memory factor.

Only one set of tests, the Repeated Span tests, has clearly demonstrated a need for both a Span Memory individual difference parameter and a Rote Memory parameter; the former accounts largely for the initial scores, and the latter appears to be more important as learning progresses. For theoretical purposes, it is fortunate that the more frequently used verbal learning tasks probably do not require a Span parameter. Such a factor might appear on tasks with lists shorter than the 10 item lists used in this experiment, but it is unlikely that it would appear when longer lists are used. Equally encouraging to the theoretical formulations is the fact that variations in method of presentation and the constancy of order have failed to demonstrate any consistent systematic effects on loadings on either the Rote Memory or Span Memory factors.

One practical implication of this study is that the selection of slow or fast learning *Ss* can be accomplished in group sessions, even in experiments that will later use individual sessions and conventional memory drum presentation. The experiment further suggests that, at least as far as individual differences are concerned, present-recall verbal learning studies may be carried out using group testing, at a saving of time and expense, and little or no loss of generality and comparability of data.

Summary

One hundred college students were given a test battery designed to investigate the loadings of a series of verbal learning tasks on the Rote Memory and Span Memory factors. The method of presentation (anticipation or present-recall), the constancy of order, and the type of task (serial or paired-associate) were varied on different tasks using consonant letters as stimuli and responses. The effects of practice over six trials and of the free recall mode of response upon the loadings on these two factors were also investigated. A set of reference factors and a set of specific factors were extracted by factor analysis. Rotations to simple struc-

ture were made within each set, but the two sets were kept orthogonal to each other.

The Rote and Span Memory factors were clearly defined as specified. The experimental verbal learning tasks displayed substantial loadings on the Rote Memory factor, and negligible or low loadings on the Span factor. Variations in method of presentation and constancy of order had no consistent systematic effect on loadings on either factor. The only substantial change of factorial composition attributable to practice was a shift of the Repeated Span tests from the Span factor toward the Rote Memory factor. The free recall mode of response had no observable effect on loadings on these factors.

Two factors (III and IV) specific to the experimental tasks were rather clearly indicated and a possible third specific factor (V) was extracted. Factor III was interpreted as a "Letter Pairs" content factor. The interpretation of Factor IV depends upon whether Factor V is retained or rejected. If it is retained, Factor IV is suggestive of an "Anticipation Method" factor; if Factor V is rejected, Factor IV's interpretation is not clear. No interpretation was offered for Factor V, if retained.

REFERENCES

BECHTOLDT, H. P. *A factor analysis of test-retest data.* Paper read at Midwestern Psychological Association, St. Louis, Missouri, May 1956.

BRENER, R. An experimental investigation of memory span. *J. exp. Psychol.*, 1940, **26**, 467–482.

CHRISTAL, R. E. Factor analytic study of visual memory. *Psychol. Monogr.*, 1958, **72**, (13, Whole No. 466).

FLEISHMAN, E. A. A comparative study of aptitude patterns in unskilled and skilled psychomotor performances. *J. appl. Psychol.*, 1957, **4**, 263–272.

FLEISHMAN, E. A. Abilities at different stages of practice in rotary pursuit performance. *J. exp. Psychol.*, 1960, **60**, 126–171.

FLEISHMAN, E. A., & HEMPEL, W. E., JR. Changes in factor structure of a complex psychomotor task as a function of practice. *Psychometrika*, 1954, **19**, 239–252.

FLEISHMAN, E. A., & HEMPEL, W. E., JR. The relation between abilities and improvement with practice in a visual discrimination reaction task. *J. exp. Psychol.*, 1955, **49**, 301–310.

FRENCH, J. W. The description of aptitude and achievement tests in terms of rotated factors. *Psychometric Monograph, No. 5.* Chicago: Univer. Chicago Press, 1951.

KELLEY, H. P. *A factor analysis of memory ability.* Princeton: Princeton University and Educational Testing Service, April 1954.

THOMPSON, G. *The factorial analysis of human ability.* (5th ed.) New York: Houghton Mifflin, 1951.

THURSTONE, L. L. *Multiple factor analysis.* Chicago: Univer. Chicago Press, 1947.

WOODROW, H. The common factors in fifty-two mental tests. *Psychometrika*, 1939, **4**, 99–108.

(Received October 31, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 12–18

A STUDY OF THE EFFECTS OF VERBALIZATION ON PROBLEM SOLVING[1]

ROBERT M. GAGNÉ AND ERNEST C. SMITH, JR.

Princeton University

Scattered throughout the literature on problem solving are occasional studies which are interpreted as indicating that acts of verbalizing during problem solving result in lessened problem solving effectiveness. While most investigators of such behavior may be inclined to expect a facilitating effect of "transfer of principles" to a final performance, studies have sometimes cast doubt upon the generality of such a finding, particularly in those instances where the principles are stated verbally.

Katona (1940), for example, found that a method which involved teaching verbal principles in solving matchstick problems to be less effective than a method of teaching by example. In a more recent study, Haslerud and Meyers (1958) found an experimental treatment in which verbally-stated principles of solution of cryptograms were given to *S*s to be less effective for solution of new cryptograms than was a treatment in which *S*s were required to discover solutions for themselves. Other findings supportive of this sort of conclusion are cited by Haslerud and Meyers (Hendrix, 1947) which suggest the superiority for transfer of "not-verbalizing" vs. "verbalizing" by *S*s themselves, in solving mathematical problems. In another study using matchstick problems, Corman (1957) failed to find significant differences in performance among groups given various amounts and kinds of verbal instructions.

Results like these contrast markedly with those of an older study by Ewert and Lambert (1932), which used as a problem the task of transferring discs graduated in size from one circle to another in a triangular configuration of three circles. In successive tasks, different numbers of discs (from three through eight) are placed in Circle No. 1, arranged in order of size with the largest at the bottom. The problem is to transfer the discs from Circle 1 to Circle 2 in the least possible number of moves, moving one at a time. One of the experimental groups in this study was given only instructions about the rules of the game; a second group was encouraged to try to find a general principle for solution; a third was given a verbally-stated principle of solution; while a fourth was given a principle plus a demonstration of the correct method with three discs. A large difference was found in the performance of groups who were given a verbally-stated principle of solution and the performance of groups who were not.

The three-circle problem is fairly difficult, and one can spend much time and many moves before discovering a principle which is truly general. Perhaps the reason for the striking contrast between the results of Ewert and Lambert (1932) and those of more modern investigators resides simply in this fact. To a participant in the problem, it seems natural indeed that his performance should improve once he knows the principle—improve not only on the problem he has attempted, but on others of a similar type. It will perhaps take more than one additional study to explicate these contrasting results.

It was our intention in the present experiment to make a further exploration into the effects of verbalizing on problem solving performance. We were interested particularly in the kind of verbalizing done by *S* himself during attempts to solve the problem, rather than by *E*.

Presumably, verbal principles provided in instructions must be repeated (perhaps to himself) by *S*, if

[1] This study was supported in part by funds granted by the Carnegie Corporation of New York. The opinions expressed are those of the authors, and do not necessarily reflect the views of that Corporation.

Data of the experiment were collected by the junior author under the direction and monitoring of the senior author. Portions of the data and results were described by Ernest Smith in his senior thesis.

they are to be effective. If we let *S* discover his own principles, in his own words, but require that he verbalize them, will this facilitate or interfere with problem solving? Marks (1951), for example, found no significant effects of providing *S*s with a typed list of principles ("elements of the problem"), but a high correlation (.83) between performance and vocalization by *S*s during solution. We were also interested in seeing whether we could establish the differences in performance suggested (but not confirmed) by Ewert and Lambert's (1932) results, of the effects of instructions to find and formulate verbally a general principle.

We chose to investigate these questions by measuring performance on a standard series of three-circle tasks of the sort employed by Ewert and Lambert, transfer to a final six-disc task of this type, and the adequacy with which *S*s could make verbal formulations of general principles. Specifically, the experiment compared the performance of groups of *S*s who solved two-, three-, four-, and five-disc problems successively, under four conditions representing combinations of two treatment variables: (*a*) a requirement to state verbally a reason for *each move* at the time it was made; and (*b*) instructions to search for a general principle which could be stated verbally after the tasks were solved.

Method

Materials.—The three-circle problem described by Ewert and Lambert (1932) was represented by three circles of 5 in. in diameter, drawn on a piece of stiff white paper, with their centers at the apexes of an equilateral triangle of side 7 in., and labeled A, B, and C. The discs were made of $\frac{3}{32}$ in.-aluminum, numbered 1 through 6, and graduated in diameter from $\frac{3}{4}$ in. to 2 in. A set of discs (two to six) is placed in Circle A, graduated with the largest at the bottom, and the problem is to move them all to Circle B so that they will be in the same order, in the smallest possible number of moves. Only one disc at a time may be moved, and it is never permitted to put a larger diameter disc on top of one with a smaller diameter. Under these rules, the fewest number of moves required (with any number n of discs) is $2^n - 1$.

The kind of principle which reduces this problem to a routine is the following: "If the number of discs is odd, move first to the circle to which you want to go eventually; if even, move first away from this circle. Continue by moving discs with odd numbers always in a clockwise direction, and the discs with even numbers always in a counterclockwise direction." There are, however, a number of other ways of formulating the second part of this principle, which are equally effective, although requiring more words.

Subjects.—The *S*s were 28 boys in Grades 9 and 10, who were assigned randomly to four experimental groups. Their ages were 14–15 yr., and their IQs were all above 110. They had volunteered to participate in studies of learning, but were paid an amount equivalent to prevailing rates of odd-job work. Each *S* was questioned closely to determine that he had no previous acquaintance with the problem, and was not used in the experiment if he had.

Procedure.—First, each *S* was shown the materials and given instructions about the rules of the game. The three-disc and the four-disc problem were then administered in succession, and each was carried to final solution (i.e., getting all discs in Circle B in the proper order). The *S* was told in each case what the minimal number of moves was. If he decided he had made a wrong move, he was permitted to go back to an earlier point in the solution, or to the beginning. The *E* made a count of all moves. The purpose of this exercise was to give all *S*s equal acquaintance with the problem, and also to provide data on equivalence of the groups.

Following this initial test, each *S* was assigned randomly to one of four conditions, each containing 7 *S*s, as follows:

Group V-SS (Verbalizing, Solution Set) was instructed to state aloud why they were making each individual move at the time they made it. In addition, these *S*s were instructed to try to think of a general rule by means of which they could tell someone

how to solve these problems, which was to be solicited afterwards by *E*. Group V (Verbalizing, No Solution Set) was required to verbalize a reason for each move, but was not instructed to try to formulate a general rule for solution. Group SS (No Verbalizing, Solution Set) was not required to verbalize, but was instructed to try to formulate a rule. Group No (No Verbalizing, No Solution Set) was simply told of the problem to be presented and its ground rules, with no additional instructions.

With these instructions, *S*s were kept at the task until they achieved a single final solution for two, three, four, and five discs in succession. One-minute rest periods were interposed between successive tasks. Number of moves was counted in each case. Of course, the making of moves was slower for those *S*s who were required to verbalize, usually to the point of slight annoyance. Following this learning session, after approximately 3 min. rest, all *S*s were presented the six-disc task as a final test. They were told that their time to solution would be measured (with a stop watch), as well as the number of moves taken. No verbalizing was required.

Recording.—For the initial test (using three and four discs), *E* simply kept a record of the number of moves made by each *S*.

During administration of the main experimental treatments (two, three, four, and five discs) number of moves was recorded. For those groups instructed to verbalize each move, *E* recorded on a prepared record sheet a brief phrase indicating the verbal statement made by each *S*. No attempt was made to make exact verbatim reproductions of what *S* said. Whenever an unusual reason was given, however, *E* attempted to record its meaning fully.

On the final test, *E* recorded the number of moves and the time required to achieve solution. After the final test was concluded, each *S* was asked to give a rule or rules for doing the problem with any number of discs, as if he were telling it to a person ignorant of the solution. The *E* recorded these verbally stated principles.

Results

Initial Test

The means and *SD*s for number of moves made by each of the four groups on the initial test composed of tasks with three and four discs were as follows: V-SS: $M = 26.4$, $SD = 4.5$; V: $M = 28.6$, $SD = 3.7$; SS: $M = 26.6$, $SD = 3.6$; No: $M = 27.9$, $SD = 4.9$. These results appeared to insure an acceptable degree of comparability ($F = .70$, $df = 3/24$; $P > .05$).

Practice Session

Mean performance curves for the four groups of the experiment are shown in Fig. 1. These depict mean number of moves in excess of the minimum for the problems with two, three, four, and five discs administered in sequence. The same measure on the final test is also shown for each group as a terminal point.

It may readily be seen that a difference among the groups, associated with the Verbalization variable, began to appear as early as the three-disc practice, and showed itself as an ever-widening difference thereafter. On the other hand, no differences of consequence appear

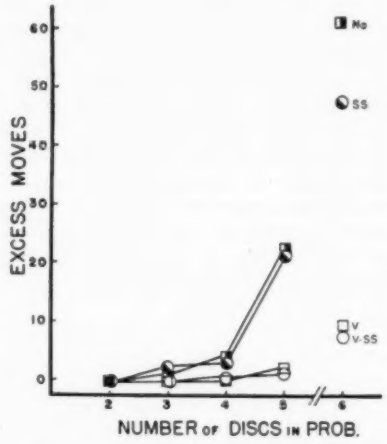

FIG. 1. Performance curves showing moves in excess of minimum required for successively administered problems of two, three, four, and five discs for the four groups of the experiment. (Values are shown for the final six-disc problem as a terminal point.)

in the performances of groups differentially treated with respect to a Solution Set, i.e., being told to look for a principle to be stated verbally. These apparent trends of the data were confirmed by an analysis of variance of the data for Trial 5. Since heterogeneity of variance was found in the raw data, they were transformed to logarithms. In this state, the null hypothesis regarding differences in variances could be rejected. An analysis of variance on these transformed data indicated significant treatment effects ($F = 6.67$, $df = 3/18$; $P < .01$). When t tests were applied to the pairs of differences between means these were found to be significant at less than the .01 level for each of the verbalization groups vs. each of the nonverbalization groups.

In contrast, neither of the pairs of groups differentiated by the presence or absence of the SS instructions yield means which are significantly different from each other. Verbalization made a significant difference, Solution Set instructions made none.

Verbalization during practice.—The verbalization done by *S*s who were instructed to give a reason for every move tended to fall into certain standard categories, with rare exceptions. There were those which were oriented toward *single* moves, like (*a*) "only possible move;" (*b*) "just to try it;" or (*c*) "don't know" (this one was quite infrequent). Then there were those which anticipated to the extent of *two* moves, such as (*a*) "to get at the larger disc;" (*b*) "to free up a space." There were also instances of reasons which apparently anticipated *sequences* of moves. One of the relatively frequent ones was (*a*) "move as with a three-disc sequence;" also in this category should perhaps come (*b*) "if disc is odd-numbered, move to circle B," and (*c*) "if disc is even-numbered, move to circle C." Finally there were some expressions of truly *general* principles, such as (*a*) "move odd-numbered discs in the clockwise direction," and (*b*) "move even-numbered discs in the counter-clockwise direction."

On the whole, the content of these "reasons" was by no means startling. It was not particularly surprising, considering the results of earlier investigators of thinking, that this form of thinking out loud was not very revealing about the nature of internal processes. A close examination of these data did not enable us to invent a method of relating specific types of verbal response to specific stages of the problem, except for the fact that the more general principles tended to occur later in practice than did the less general ones.

Final Task Performance

Mean number of moves in excess of minimum, and mean times (min.) to solution, for each of the four groups of the experiment on the final six-disc task are shown in Table 1. Here it can be seen that under standard conditions for all the groups, the differences in moves which appeared on the five-disc task are even greater on the six-disc task. Again the variances are markedly different between the verbalization and nonverbalization pairs of groups. A logarithmic transformation was found to make possible the assumption of homogeneity of variance. An analysis of variance performed on these transformed data confirmed significant treatment effects ($F = 9.13$, $df = 3/18$; $P < .01$). Application of t tests to differences between individual means indicated significance at the .01 level for comparisons of the verbalization and nonverbalization groups. Dif-

TABLE 1

MEANS AND *SD*s FOR NUMBER OF MOVES IN EXCESS OF MINIMUM, AND FOR TIME OF PERFORMANCE, ON THE FINAL SIX-DISC TASK FOR THE FOUR GROUPS

Measure	Groups			
	V-SS ($N=7$)	V ($N=7$)	SS ($N=7$)	No ($N=7$)
Moves in excess of minimum				
Mean	7.9	9.3	48.1	61.7
SD	8.6	7.9	34.8	40.2
Time (min.)				
Mean	4.2	3.8	10.1	10.0
SD	2.2	2.0	4.1	5.5

ferences between means for Groups SS and No, and Groups V-SS and V, are not significant.

An entirely similar set of comparisons exists with the means of time to solve, shown in Table 1. Here the F test for the total data is significant at lower than the .05 level ($F = 4.90$, $df = 3/18$). The t test indicates significance of differences of means associated with the contrast of verbalization vs. no verbalization (V-SS vs. SS; V-SS vs. No; V vs. SS; V vs. No) at better than the .05 level. Other differences in means are not significant.

Stating the Verbal Principle

Records of statements made by *S*s in response to the instruction to state verbally how to solve these problems were examined by the 2 *E*s independently, and rated "Inadequate," "Partial," or "Complete." The following agreements were made beforehand: "Inadequate" was to mean any principle which was incorrect, irrelevant, or so fragmentary that it would be of no aid to anyone in finding solutions. "Partial" meant one of the two major parts of the principle, concerned with starting (move odd-numbered disc to the goal circle, etc.) or with procedure during solution (even numbers move clockwise, etc.) "Complete" meant both of these major parts.

With these criteria, it was quite easy to classify the principles the *S*s stated, and there were no differences in judgment of the 2 *E*s. These results are shown in Table 2.

Considering first the dichotomy Inadequate vs. Partial or Complete, the Fisher exact probability test was applied to comparisons of the experimental treatments. The probability of the observed contrasts between V-SS and No, and between V and No were found to be .002; between SS and No, .014. All other comparisons yielded high probability values. So far as being able to state even *partial* verbal principles, Group No is significantly poorer than those who verbalized; for those who did not verbalize, there is a significant effect of instructions to look for a general principle. Second, the dichotomy Inadequate or Partial vs. Complete was considered by this same method. Here the only low probabilities were .036 between Groups V-SS and SS and also Groups V-SS and No. The group required to verbalize as well as to try to formulate a verbal principle,

TABLE 2

NUMBER OF INADEQUATE, PARTIAL, AND COMPLETE VERBAL PRINCIPLES STATED BY *S*s IN EACH GROUP

Group	*N*	Number of Instances of Principles Judged as:		
		Inadequate	Partial	Complete
V-SS	7	0	3	4
V	7	0	5	2
SS	7	1	6	0
No	7	6	1	0

was thus found to be superior to Groups SS and No in the formulation of *complete* verbal principles. It would appear that verbalizing during practice is the most important factor accounting for this difference.

Discussion

Our results indicate that requiring individuals to verbalize while practicing the three-circle problem is a condition which is significantly related to superior performance in problem solving, which begins to show itself during practice with two-, three-, four-, and five-disc tasks, and is maintained on a final six-disc task. There is a strong suggestion also, that verbalizing is the most important factor at work in producing a greater number of individuals who can, at the close of practice, state fully adequate verbal principles of task solution.

Instructions to try to formulate verbal principles for solution of these tasks appear to have no effect on performance of the task, and also an insignificant effect (by themselves) in producing a greater frequency of fully adequate principles. They do, however, produce more verbal principles which are at least partially adequate, than is the case when they are omitted. It will be apparent in interpreting these results that we cannot be certain how great a contrast in "set to look for solution" was actually achieved by the instructions used. After all, the noninstructed groups as well as the instructed ones underwent an initial period of testing on the three- and four-disc tasks, and may have generated in themselves a set to look for a general solution.

Surely the most striking finding is the effect of verbalization by the *Ss* during practice. Such verbalization is of course very different from that which may result from providing *Ss* with verbally stated principles of solution, as was done in the studies of Ewert and Lambert (1932) and Haslerud and Meyers (1958). Whatever principles were used by the individuals in the present experiment, they must have been discovered by the *Ss* themselves. In this respect, at least, the present findings are not inconsistent with those of Haslerud and Meyers, nor with other results indicating the effectiveness of self-discovered principles for problem solving (e.g., Gagné & Brown, 1961).

It is perhaps of some importance to emphasize that the individuals who were required to verbalize also took more time to make successive moves. We cannot tell from these data, of course, whether this more deliberate pacing of the task may have had some effects on the performance. The additional time between moves, however, was "filled" time, taken up entirely with the act of verbalization. Thus it would not be reasonable to suppose that the verbalization groups had any greater opportunity than did the nonverbalization groups for the deliberate rehearsal and recall of successful and unsuccessful moves. Verbalization is the most obvious variable at work; if other events are contributing to the results because of their association with time differences, it is not immediately apparent what they are.

As we pointed out, the content of the verbalizing during practice was fairly pedestrian and to some extent routine, so that it could be readily categorized. What then accounts for its effect on problem solving? In answering this question, we have no theory to call upon. It would appear that requiring verbalization somehow "forced the *Ss* to think." In other words, this treatment may have had the effect of constantly prodding the *Ss* to think of new reasons for their moves, particularly since they may have gotten a little tired of giving "the same old reasons" over again. This conception of the treatment would assign it a role similar to that of instructions to think of new word associates, as in the studies of Maltzman, Simon, Raskin, and Licht (1960). But this is a speculation which obviously does not come directly from present data. It would, however, be an inter-

esting route for follow-up studies to take; namely, testing whether instructions to "think of new reasons," to be stated verbally, would have a facilitative effect on the solution of a task like that used here.

Summary

A study was conducted to determine the effects of (*a*) requiring *Ss* to verbalize during practice, and (*b*) instructions to find a general principle to be stated verbally, on problem solving performance. The three-circle task described by Ewert and Lambert (1932) was used. The *Ss* were 28 ninth and tenth grade boys, divided randomly into four groups of 7 each. Following the initial test on tasks with three and four discs, *Ss* were assigned to four groups given different treatments representing combinations of the two experimental variables. The practice session consisted of two-, three-, four-, and five-disc tasks administered successively. Records were kept of the verbalizing, for those groups who engaged in it, and of number of moves to achieve solution, for all groups. Following this, a final task using six discs was administered to all *Ss* under standard conditions, and a record was made of number of moves as well as time to solution.

Significant differences were found between the scores of those groups who were required to verbalize and those who were not, in terms of number of moves. Similar differences based on time scores were found in the contrast between verbalization and nonverbalization groups. Differences between other pairs of groups were not significant.

Comparisons of success in stating verbal principles applicable to the solution of three-circle problems were made among judged categories of Inadequate, Partial, and Complete, for all groups. A significant difference was obtained between verbalizing and nonverbalizing groups, and one which favored the "instructed" groups when the dichotomy Inadequate or Partial vs. Complete was tested.

The results appear to indicate that requiring *Ss* to verbalize during practice has the effect of making them think of new reasons for their moves, and thus facilitates both the discovery of general principles and their employment in solving successive problems.

References

Corman, B. R. The effect of varying amounts and kinds of information as guidance in problem solving. *Psychol. Monogr.*, 1957, **71**(2, Whole No. 431).

Ewert, P. H., & Lambert, J. F. Part II: The effect of verbal instructions upon the formation of a concept. *J. gen. Psychol.*, 1932, **6**, 400–413.

Gagné, R. M., & Brown, L. T. Some factors in the programing of conceptual learning. *J. exp. Psychol.*, 1961, **62**, 313–321.

Haslerud, G. M., & Meyers, S. The transfer value of given and individually derived principles. *J. educ. Psychol.*, 1958, **49**, 293–298.

Hendrix, G. A new clue to transfer of training. *Elem. sch. J.*, 1947, **48**, 197–208.

Katona, G. *Organizing and memorizing.* New York: Columbia Univer. Press, 1940.

Maltzman, I., Simon, S., Raskin, D., & Licht, L. Experimental studies in the training of originality. *Psychol. Monogr.*, 1960, **74**(6, Whole No. 493).

Marks, M. R. Problem solving as a function of the situation. *J. exp. Psychol.*, 1951, **41**, 74–80.

(Early publication received July 24, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 19–22

COUNTERTRAINING OF A SIMPLE SKILL WITH IMMEDIATE AND 1-WEEK DELAYS OF INFORMATIVE FEEDBACK[1]

FRANCIS J. RYAN[2] AND EDWARD A. BILODEAU

Tulane University

Delaying knowledge of results (KR) or informative feedback by 7 days has the effect of interpolating a 7-day layover between successive responses. Such spacing retards the learning of a simple, discrete motor response. The effect of delay of KR is another question, as was made clear in a series of five experiments by Bilodeau and Bilodeau (1958). They found no effects attributable to the time between R and KR, and a negative effect attributable to the time between one R and the next. The interresponse intervals and delays of KR ranged from 20 sec. to 7 days. The interval between KR and the next R (post-KR interval) also ranged up to 7 days.

Their Exp. 4 and 5 are particularly relevant here. Experiment 4 used a 1-day interval between Rs, and showed a group with KR delayed by 1 day to be superior to a group with no delay. The most plausible hypothesis was that delay within a relatively massed series of Rs was unimportant while to delay KR would be best with long interresponse intervals, i.e., delay of KR interacts with interresponse interval. Experiment 5 was designed for confirmation. Two groups employed a 7-day interval between Rs, one was given KR immediately; the other was delayed for 7 days. The results were disappointing, though not unexpected. Neither group learned more than a trivial amount, so there was no opportunity for a difference as a function of delay. The wide spacing of Rs had again shown its powerful effect. Experiment 5 had too few *S*s and trials. Both were in short supply since R was scheduled at the rate of one per week. The present experiment used a greater number of *S*s and Rs than did Exp. 5.

Still another consequence of the week-long interval was the large between-*S* variance; that is, a great deal of hunting behavior was shown. The problem of the unwieldy variances led to a review of techniques for regulating learning in human beings. No existing task for studying the learning process in human *S*s was judged suitable. For example, too often, conditioning takes place too fast or not at all. Control over extinction rates poses even greater problems. Further, in the usual task, control over the variances is left to luck.

Some control over between-*S* variance is possible. The present solution is to train *S*s to homogeneous performance. After *S*s are performing similarly, the experimental treatments are introduced at the same time that the task is changed to require new Rs instead of the previously trained R.

The writers wish to introduce a practical countertraining technique to the motor skills area and propose that informative feedback is the natural vehicle toward this end. For the training period, *S*'s feedback error will be computed with regard to a particular target. After training is completed, the target is shifted systematically, the only indication of change lying in KR. To keep KR errors to a minimum, *S* must learn to modify successive Rs. The experimental variable—delay of KR—is introduced simultaneously with the target changes.

[1] This work was supported by a grant from the National Science Foundation to E. A. Bilodeau, principal investigator. It also represents an abridged version of a thesis submitted to the Graduate School of Tulane University in partial fulfillment of the requirements for the Master's Degree.

[2] Now at the State University of Iowa.

METHOD

Subjects.—Seventy students at Tulane University volunteered for eight sessions at weekly intervals. They were assigned to two groups and run individually according to a counterbalanced schedule. To assure better attendance, *S*s completing the 8 weeks were promised $3.00; the 6 best performers expected prizes ranging from $10.00 to $50.00. Nine *S*s failed to keep appointments. Also voided was an injured *S*, 1 misrun by *E*, and 2 who failed to meet a pre-established training criterion.

Apparatus.—Manual Lever C, a modification of an earlier device developed by Bilodeau and Ferguson (1953) was used. It is a stick or lever mounted inside a 23 × 17 × 17 in. wooden cabinet. Lever displacement (*S*'s response) is measured by means of a scale on the back of the cabinet. The *S* was unable to see the lever and scale.

The task was to move the lever from a zero (starting) position to a second position (target). The *S* attempted to hit the target on each move. A response 1° short of the target was reported as "short by 5 units," a move 4° beyond the target was reported as "long by 20 units," etc. Units were not defined for *S*.

Procedure.—On Day 1 (training), 8 Rs were required. The first 7 of these were all followed by immediate KR. The experimental variable was introduced after R8 when half of the *S*s were switched to a 7-day delay schedule and the remaining *S*s continued with immediate KR. R9 was required on Week 2, R10 on Week 3, and so on through Week 8. For purposes of reference: the first 8 Rs have been assigned to training; the countertraining procedures begin immediately after the termination of R8, and the countertraining effects begin (potentially) with R9 (Week 2).

Three substantial changes in procedure followed R8: (*a*) the interresponse interval was lengthened to 7 days; (*b*) the temporal locus of KR was changed, a week being interpolated between R and KR for Group D, and a week being interpolated between KR and R for Group I; and (*c*) the target position was decreased by 4° per trial. KR was delayed by a trifle (2 or 3 sec.) for Group I; for Group D, the post-KR period was held close to 15 sec. These temporal intervals minimize the confounding problem (Bilodeau & Bilodeau, 1958), since the KR delay is a small fraction of the intertrial interval for Group I, and the 15-sec. post-KR interval is also a small fraction of the intertrial interval for Group D. Elaborate precautions were taken to withhold any cue of the target changes other than that provided by KR.

Each *S* sat facing the lever for about 15 sec. during a demonstration of standard posture, grip, and rate of movement. Next, *S* was coached to avoid moving too fast (reduces kinesthetic cues) and moving too slow (produces chatter). The task was described and *E* stated that KR would assist in learning it. A "practice" move of "about half an inch" was then called for (R1) and immediate KR was given. R2 through R8 followed at 20-sec. intervals.

Following R8, KR was withheld from Group D, both groups waited 15 sec. (as if for another R) and were dismissed. The next week, *E* reminded *S* of the task, posture, rate of movement, and then gave KR to Group D. All *S*s waited 15 sec. before making R9. After responding, *S*s of Group I were given immediate KR and all *S*s waited 15 sec. before dismissal.[2]

All *S*s were taught to make a movement from 26° below the vertical to a position 10° beyond the vertical, an arc of 6.9 in. Using the preferred hand, half the *S*s moved from left to right, and half from right to left. If a delayed *S* asked for KR after responding, or if an immediate *S* sought a reminder of KR just before executing a response (such questions were rare), he was told that *E* lacked time to figure out the score right now, or that *E* had forgotten.

RESULTS

Figure 1 shows the mean amplitude of R for all 15 trials. The corresponding *SD*s have been placed in Table 1. Both Groups I and D learned the training task fairly quickly as is shown by the increasing means and decreasing *SD*s. The eight trials of training were just about sufficient to produce maximum accuracy (means) and near minimum variability (*SD*s). Further, Groups I and D, having been trained alike, behaved quite alike.

[2] After R15 of Week 8, both groups were given immediate KR, held 15 sec., and dismissed. One hundred and sixty days later the *S*s were brought back for a test of recall. The recall-test data were not part of the thesis work and are not reported here.

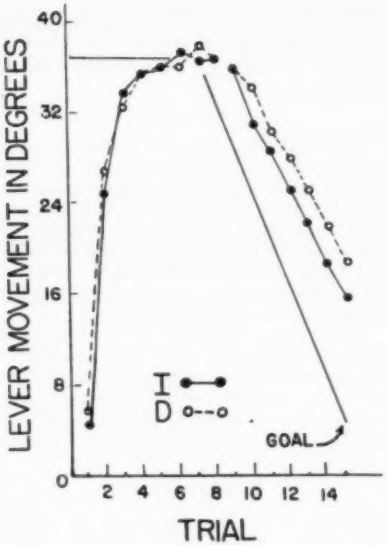

FIG. 1. Mean response amplitude for training (Trials 1–8) and for countertraining (Trials 9–15) for Groups I and D.

The means of the countertraining trials (9–15) show both groups learning to reduce the amplitude of R with successive trials; the mean square for trials was significant ($P < .01$). Though R was at the rate of one per week, there can be no doubt of a powerful learning effect in the means of both groups.

The solid line in Fig. 1 represents the loci of the targets. Comparing the location of Curves I and D to that of the target, we see that both groups lagged behind with the greater lag shown by Group D ($F = 5.01$; $df = 1/55$; $P = .05$). There is also a fairly constant difference between I and D; the interaction term produced an F less than 1.0. Group I has less than a one-trial advantage over D. This advantage first appears on Trial 2 of countertraining and does not increase thereafter. The differences between the means of Groups D and I (D − I) on Trials 9–15 were −0.2, 3.7, 1.4, 2.7, 2.7, 3.0, and 3.4. The overall difference is 2.4.

TABLE 1

*SD*s OF RESPONSE AMPLITUDES FOR GROUPS I AND D

Trial	I	D	Trial	I	D
1	2.1	3.9	9	7.9	6.4
2	18.2	14.5	10	6.9	5.3
3	12.5	10.7	11	5.1	4.5
4	9.2	8.0	12	5.1	5.2
5	5.6	6.1	13	5.2	5.5
6	8.6	5.1	14	6.0	6.0
7	3.4	3.6	15	8.1	10.9
8	3.3	2.9			

Examination of the *SD*s in Table 1 shows similar variabilities from I to D. The major effect occurs between Trials 8 and 9 in that both groups become more variable as the intertrial interval was lengthened from 20 sec. to 7 days. There is no special effect at Trial 10 as there was with the means. During the last 2 weeks a few *S*s began to perform erratically.

The mean absolute errors were also examined carefully, but they are not reproduced here because the findings were highly similar to those already reported. Suffice it to say that the target moved 4° per week whereas the *S*s modified R by about 2° per week, the Group I *S*s showing a very small advantage on Trial 10.

Comments from *S* were not solicited. However, 6 of the 27 *S*s in Group D and 4 of the 30 *S*s in Group I made some remark about the moving goal. As in previous studies (Bilodeau & Ryan, 1960), almost no one complained about, or remarked on, the delay of KR.

DISCUSSION

This experiment was designed to provide further evidence on an earlier suggestion that the damage done by very long intertrial intervals may be lessened

by delaying KR. No support for this idea was found.

The difference between Groups I and D at the .05 level and the insignificant interaction of I and D over trials were the two major findings. In five earlier studies there was no significant superiority of I over D, and one of D over I. Significance beyond the .01 level should be and was required for rejecting the null hypothesis; hence the present .05 difference is ascribed to error and chance. In addition to the previous studies, there were also internal reasons for the .01 criterion: (*a*) If delay of KR hinders the learning of this skill, the difference between Curves I and D should increase with trials. There was no evidence for such an interaction. (*b*) The difference between I and D did not originate with the first week of delay. Those who choose to reject the null hypothesis must propose a delayed effect. (*c*) Statements volunteered by *S*s contained very few references to KR—its delay and changing nature.

Bilodeau and Bilodeau (1958) have reported that a group with a 1-day intertrial interval and a 1-day delay of KR was superior to the group with no delay and a 1-day intertrial interval. The present experiment is similar, but the result is not. The treatment differs in that the intertrial interval is longer and the task superficially different. The previous finding has been placed under a cloud by the data reported here; the present finding, of course, is equally dubious in view of the earlier contradiction.

Summary

Fifty-seven *S*s were first taught in eight massed trials to displace a lever by a certain amount. Following the eighth response, the intertrial interval became 7 days, the amplitude of response required was changed in order to obtain countertraining, and the delay in KR was changed for half the *S*s. For one group KR occurred immediately; for the other KR was withheld for 7 days.

Both groups learned to modify their responses reasonably soon. The immediate-KR group outperformed the delay group ($P = .05$), but the null hypothesis was not rejected because several other internal and external results contradicted the .05 outcome.

References

Bilodeau, E. A., & Bilodeau, I. McD. Variation of temporal intervals among critical events in five studies of knowledge of results. *J. exp. Psychol.*, 1958, **55**, 603–612.

Bilodeau, E. A., & Ferguson, T. G. A device for presenting knowledge of results as a variable function of the magnitude of the response. *Amer. J. Psychol.*, 1953, **66**, 483–487.

Bilodeau, E. A., & Ryan, F. J. A test for interaction of delay of knowledge of results and two types of interpolated activity. *J. exp. Psychol.*, 1960, **59**, 414–419.

(Received September 1, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 23–28

A TEST OF THE RESPONSE PROBABILITY THEORY OF PERCEPTUAL DEFENSE[1]

MICHAEL J. GOLDSTEIN

University of California, Los Angeles

Two recent reviews on the influence of motivation upon perception (Eriksen, 1958; Goldiamond, 1958) have concluded that the dilemma posed by phenomena of perceptual defense is more apparent than real. The finding of lower thresholds for anxiety-arousing words (when compared to thresholds for neutral words) would appear to imply the existence of a supersensitive scanning mechanism which first scans visual stimuli and then inhibits the full recognition of images likely to arouse anxiety. Eriksen (1958) and Goldiamond (1958) argue that another, more parsimonious explanation can account for the phenomena of perceptual defense, an explanation which emphasizes the operation of response variables. These authors suggest that as a consequence of prior conditionings, *S* develops differential probabilities of using various words as responses in a perceptual situation. In the case of perceptual defense, responses which have become conditioned stimuli for anxiety will be less likely to be evoked under conditions of marginal perceptual stimulation than other stimuli of similar physical contour which have not shared this association with noxious stimulation. This lowered probability of evocation (negative response bias) of anxiety-arousing words can account for the perceptual defense effect.

If *S* is going to call a particular class of words less frequently than another, then the possibility of a chance correspondence between his response and the correct stimulus according to *E*'s answer sheet is less likely for the undercalled class than for the overcalled class of words. The higher threshold for anxiety-arousing words thus becomes a simple artifact of the operation of response factors and does not require the postulation of a prescanning mechanism in the visual system.

This response probability theory emphasizes the necessity of separating two sources of variance in perceptual recognition scores, one component which appears to be independent of the presence of the perceptual stimulus and a second source which appears to require the presence of the perceptual stimulus before certain effects can be noted. Goldiamond (1958) refers to these two components as extraneous variance and perceptual variance and argues for the necessity of isolating them. The purpose of the present report is to attempt an experimental analysis of the phenomena of perceptual defense in an effort to isolate these two sources of variance. In attempting this analysis, four basic questions are asked: (*a*) Does a negative response bias exist for anxiety-arousing words when no discriminative stimulus is presented? (*b*) If present, is such a

[1] This study was supported by a grant from the Faculty Research Committee, University of California, Los Angeles. An abridged version of this paper was presented at the Western Psychological Association meetings, April 1959 in San Diego, California. The author would like to express his appreciation to Samuel Himmelfarb for running *S*s and analyzing data and to Norman Anderson for his generous help on statistical matters.

bias of sufficient magnitude to account for the differences in accuracy, when no discriminative stimulus is presented? (*c*) What changes in response bias occur when a discriminative stimulus is present, and are they sufficient to account for all of the variance in accuracy scores? (*d*) What effect does increasing the amount of available information have upon response bias and accuracy?

METHOD

Subjects.—There were 60 college student *Ss*, 30 of each sex. Each *S* was assigned at random to one of three experimental groups: a Stimulus Absent (SA) group, a Stimulus Present (SP) group, and an Increasing Information (II) group, with the restriction that there were 20 *Ss* in each group.

Selection of anxiety-arousing and neutral stimuli.—In a preliminary session each *S* was administered a 117 item word association test. The stimuli in this test were constructed in triads, each word designed to be potentially anxiety arousing being paired with two other words of the same frequency of usage, same length, same starting letter, and neutral in content. The triads were not adjacent in the test, but were randomly interspersed throughout. The *S* was instructed to respond to the stimulus word with the first word that came to mind; his reaction time (RT) (in .1 sec.) and his verbal response were recorded. After completion of the test, the list was read a second time and *S* was instructed to give the same response that he gave on the first administration.

A word was selected as anxiety arousing for *S* if there was a prolonged RT to the word and a reproduction fault on the second administration. The RT for the anxiety word had to be several seconds longer than the RT to the two other members of its triad for a word to be considered anxiety arousing. For each anxiety-arousing word selected, a neutral word was selected from the same triad for which no block or reproduction fault was evident. Four anxiety-arousing and four neutral words were selected for each *S*. Since both classes were selected from the same triad, the final list of eight words involved four pairs of words with identical frequencies of usage, length, and starting letter.

Perceptual task.—Following the word association test, *S* was scheduled to return within 48 hr. Upon his return *S* was introduced to the Gerbrands tachistoscope and an estimate of his visual sensitivity was obtained. The *S* was shown a list of eight words and was asked to learn them. The words were all of extremely high frequency of usage and included three key words: LIGHT, DOLLAR, and RETURN. Although *S* was told that all eight words were to be flashed in the tachistoscope, only the three key words were flashed according to a prearranged random order. These key words were each presented twice at a series of increasing exposure times until a setting was reached at which *S* could identify any one of the three words on two nonconsecutive trials. This exposure time was considered a lower bound of the range of *S*'s visual sensitivity. Once this exposure time was determined, the tachistoscope was then set at 50–100 msec. below this figure.

For Groups SA and SP, the exposure time set for *S* was used for all trials of the experiment; however, Group II received the first block of 32 trials at this setting and each subsequent block of 32 exposures at a setting 50 msec. above the previous block's level.

Following the determination of sensitivity *S* was shown another list of eight words which had been selected from his word association performance and was told that now only these eight words would be flashed in the tachistoscope. A minute was given to *S* to look over the list, after which it was attached to the tachistoscope. Specifically, *S* was instructed:

> This is a study dealing with perception. Here is a list of eight words which will be presented one at a time, in a random order, in this machine. You are to look into the eyepiece and judge which of the eight stimuli is present on each trial by referring to your list. The words will be presented rapidly so that at times it will be very difficult for you to see the words and you will frequently have the feeling that you are guessing. Don't be disturbed by this, just guess on every trial.

The *Ss* in Group SA were instructed that words were to appear in the tachistoscope, but *E* actually flashed a single nonsense figure (five typewritten hashmarks) for the 128 trials of the experiment. Group SP actually received a word on each of the 128 trials according to a prearranged random order. The only limitation on the random order was that both classes of words (anxiety arousing and neutral) had to be presented equally often and that in each of the four blocks of 32

trials, each of the eight words had to appear equally often. These limitations were placed on the random order in order to provide for an analysis of a possible effect over trials upon bias and accuracy. For Group II words were flashed on each of the 128 trials just as for Group SP.

The experiment was conducted in a windowless, black, experimental room, the only illumination coming from a 7-w. bulb.

Pseudoaccuracy scores.—To answer the question of whether response bias can produce the perceptual defense effect, it was necessary to compute pseudoaccuracy scores for Group SA. Pseudoaccuracy scores were obtained by scoring *S*'s responses as correct or incorrect according to the random order of stimulus presentations actually used with the other two experimental groups. These scores would be expected to vary about a chance mean of 12½% correct.

Results

The results were analyzed in terms of performance over all 128 trials and in terms of performance in successive blocks of 32 stimulus presentations. The data for over-all performance will be presented first, as they correspond to the kind of data typically reported in perceptual defense studies. It should be made clear, however, that the results are not described in terms of thresholds, but in terms of accuracy scores over successive runs at a single stimulus presentation or over blocks of increasing exposure times. This was done to avoid some of the pitfalls judged to be present in the classical ascending method of limits, in which termination of series by a correct response undoubtedly alters the system of response biases for the remaining words in the set.

First we turn our attention to the question of whether response bias exists when no discriminative stimulus is present. The data from Group SA were analyzed in terms of the total number of emotional words called by each *S* over the 128 trials. Since 64 calls would indicate no bias, each *S*'s bias score was subtracted from 64 to yield a difference score. The mean difference score for Group SA was −4.55 (−2.3 for men, −6.8 for women), indicating a negative response bias with regard to emotional words. Analysis of variance of these difference scores revealed a significant F (6.53 for 1/18 *df*, $P < .025$) for the mean, but no significant F for sex of *S*.

The next issue was whether this degree of response bias can bring about differences in accuracy scores which would ordinarily be interpreted to be due to perceptual factors. Each *S*'s record was scored for pseudoaccuracy (i.e., the number of times he would have identified the stimulus had it actually been presented). Separate scores were obtained for anxiety-arousing and neutral words. An *S*'s mean pseudoaccuracy score for neutral words was subtracted from his pseudoaccuracy score for anxiety-arousing words, yielding a distribution of difference scores. The mean difference was −2.45 (−.3 for men, −4.6 for women), suggesting lower pseudoaccuracy scores for the anxiety-arousing words. Analysis of variance of the difference scores revealed an F of 8.87 for the mean ($P < .01$ for 1/18 *df*), indicating that the mean difference is statistically different from zero, and an F of 6.81 for sex ($P < .025$ for 1/18 *df*). The latter finding indicates greater differences in pseudoaccuracy scores for women than for men. The analysis of pseudoaccuracy indicates clearly that the negative response bias is sufficient to produce significant differences in pseudoaccuracy. Had a real perceptual stimulus been presented, we would have erroneously concluded that the poorer rate of recognition for anxiety-arousing words (compared with neutral words) was a

function of perceptual factors rather than response factors!

Effect of presenting a discriminative stimulus.—The fact that response bias can produce the results described above is no assurance that this variable is the sole one in operation when a discriminative stimulus is present. In an effort to study the possibility that other factors are operating, the data from Group SP were analyzed in terms of response bias and accuracy. Bias was computed in the manner described above. The mean difference score for Group SP was −4.90 (−5.30 for men, −4.50 for women), indicating a negative response bias, but no tendency for the bias to relate to sex of *S*. An analysis of variance of the bias difference scores revealed an F of 16.32 ($P < .001$, for 1/18 *df*) and an insignificant F for sex. This analysis indicates that there exists a significant tendency to undercall anxiety-arousing words when a discriminative stimulus is flashed in the tachistoscope.

A similar analysis was done on the accuracy scores for anxiety-arousing and neutral words. The number of stimuli in each class correctly identified by *S* over 128 trials was computed. A difference score was then computed for each *S* between his accuracy scores for anxiety-arousing words and his accuracy score for neutral words. The mean of these difference scores was −4.50 (−6.10 for men, −2.9 for women) which was statistically different from zero ($F = 12.62$, $P < .005$ for 1/18 *df*.) This merely duplicates the well known perceptual defense effect in which *S* shows poorer rates of decoding for anxiety-arousing words than for neutral words.

Comparison of stimulus Groups SA and SP.—A stated purpose of the present research was the isolation of perceptual variance from response variance. Since the data from Group SA represented a "pure" measure of differences in accuracy due to response bias, a test was made of whether the difference in accuracy found in Group SP was significantly greater than could be accounted for by response bias alone. Using a distribution of difference scores, an analysis of variance was done to determine if the mean difference in accuracy between anxiety words and neutral words in the Group SP (−4.50) was significantly greater than the mean difference in accuracy (−2.45) found in Group SA. An analysis of variance of these difference scores revealed that the variance between groups was not statistically significant ($F=1.88$, $P > .20$ for 1/38 *df*). Thus, the magnitude of the perceptual defense effect found in the presence of a discriminative stimulus is not significantly greater than the magnitude that would be expected on the basis of response bias alone.

Performance over trials.—The order of presentation of stimuli was designed to permit a four-block analysis over trials, with 32 stimulus presentations in each block. This was done to examine any trends in both call rate and accuracy as a function of repeated presentations of anxiety-arousing and neutral stimuli. Figure 1 presents the data for response bias over blocks for the three experimental groups. It can be seen that the trends over blocks is different for the groups which received stimuli than for Group SA. Both Groups SP and II exhibit an increase of negative response bias over blocks, while the data for Group SA suggest a gradual decline in bias. A statistical comparison of response bias was done on the trends for Groups SA and SP in the manner suggested by Grant (1956) for repeated measurements

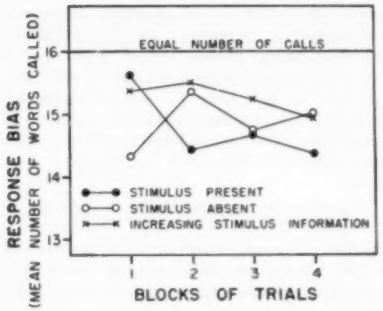

FIG. 1. Mean number of anxiety-arousing words called over the four blocks of the experiment by *Ss* in the three groups.

data. The analysis was restricted to the linear component scores. The mean linear score for Group SA was +4.1 (+0.10 for men; +4.0 for women) and the mean linear score for Group SP was −6.5 (−5.0 for men; −1.5 for women) supporting the visual impression of the two curves. However, the analysis of variance of these linear scores is not significant for groups ($F = 1.170$, $P > .10$ for 1/36 *df*), sex ($F = 1.68$, $P > .10$) or the Groups × Sex interaction ($F < 1$), all tests for 1/36 *df*.

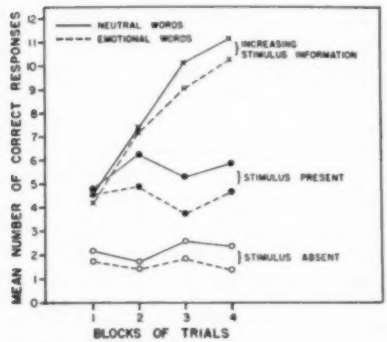

FIG. 2. Mean number of anxiety-arousing (emotional) words and neutral words correctly identified over the four blocks of the experiment by *Ss* in the three groups.

A similar analysis by blocks was done on the accuracy scores and is presented in Fig. 2. It can be seen that very little increase in accuracy scores occurred with repeated presentations of the stimulus (Group SP) and that improvement in accuracy is directly related to longer exposure times of the stimulus (Group II). Further, differences in accuracy appear between anxiety-arousing and neutral words over blocks of trials, even when the later blocks of trials imply longer exposure times. Rather than extinguishing over the course of the experiment, the perceptual defense effect appears to become enhanced with practice! A comparison of the trends in Groups SA and SP was done in the manner described above for bias scores. No significant *F*s for slope were found.

DISCUSSION

The present study represents an attempt to isolate two sources of variance in perceptual recognition scores: variance related to highly overlearned response habits and variance related to the presence of a visual stimulus. The results indicated that *S* does enter the perceptual situation with clearly defined response habits which are not under the control of the perceptual stimulus and which can influence *S*'s recognition score. With regard to the perceptual defense effect words classified as anxiety arousing have less probability of being used as recognition guesses and this negative response bias makes a correct signal detection less likely when a discriminative stimulus is presented.

Comparison of Groups SA and SP did not indicate that the presence of a perceptual stimulus produced a perceptual defense effect greater than that expected on the basis of response bias alone. Thus, the present research supports a response probability theory of perceptual defense.

It should be noted that the present

method of isolating response variance from perceptual variance does not exhaust the methods for attacking this problem. Despite the absence of a significant difference between Groups SA and SP in the magnitude of the perceptual defense effect, it is possible that different sets of variables could be producing the same effect in the two groups. More conclusive evidence for the exclusive role of response bias in producing the perceptual defense effect will come from studies which find a means for equalizing the probability of usage of the two classes of indicator responses.

Summary

The role of response factors in producing the perceptual defense effect was evaluated. Sixty *S*s were assigned to three groups: Group SA (Stimulus Absent), Group SP (Stimulus Present), Group II (Increasing Information). Stimuli, presented tachistoscopically, were pairs of words (half anxiety arousing, half neutral in meaning) selected from *S*'s word association protocol.

Results were as follows: (*a*) A negative response bias (for anxiety-arousing words) is present when no discriminative stimulus is present. (*b*) This negative response bias would be sufficient to produce significant differences in accuracy, as estimated by pseudoaccuracy scores. (*c*) Comparison of pseudoaccuracy scores from Group SA with accuracy scores from Group SP indicates that presence of a discriminative stimulus did not produce a perceptual defense effect greater than would be expected by response bias alone.

The results are interpreted as supporting a response probability theory of perceptual defense. It is suggested that final evaluation of this theory awaits studies in which perceptual recognition scores are obtained under conditions in which no response bias is possible.

References

Eriksen, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958*. Lincoln: Univer. Nebraska Press, 1958.

Goldiamond, I. Indicators of perception: I. Subliminal perception, subception, unconscious processes: An analysis in terms of psychophysical indicator methodology. *Psychol. Bull.*, 1958, **55**, 373–411.

Grant, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

(Received October 1, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 29–31

PREDICTING INSTRUMENTAL PERFORMANCE FROM THE INDEPENDENT RATES OF CONTINGENT RESPONSES IN A CHOICE SITUATION [1]

AARON J. BROWNSTEIN [2]

University of Missouri

The present experiment derives from a theory of positive reinforcement (Premack, 1959, 1961) in which the necessary and sufficient condition for reinforcement is a rate differential. That is, if a response of higher rate is made contingent upon the occurrence of a response of lower rate, then an increment in the rate of the lower rate response will obtain. The determination of higher and lower rate response is carried out in a noncontingency, sometimes called operant level, situation. (See Premack, 1959, 1961, for a more detailed discussion of this procedure.) The rate so obtained is called independent rate. Thus, the theory attempts to predict rate in the contingency situation, called dependent rate, from the independent rates.

The theory assumes that the increment given the lower rate response is proportional to the independent rate of the contingent response. Furthermore, independent rates of like numerical value are of like reinforcement value, irrespective of the responses in which the rate is manifested or the parameters with which the rate is obtained.

The theory contains no special provisions for predicting behavior in a choice situation. Dependent rates in the choice situation should be proportional to independent rates of the contingent responses, as measured *in the choice situation*. Thus the theory recognizes the choice situation only as a possible parameter of the independent rate of the contingent responses.

Within the present framework prechoice dependent rate differences would be expected to be maintained, modified, or reversed in the choice situation only as the introduction of choice maintains, modifies, or reverses the difference between the independent rates of the contingent responses. This approach differs radically from approaches in which instrumental response strength for the subsequent components of the choice are made the basis for predicting behavior in the choice situation.

Young and Greene (1953) have shown that, when *S*s are permitted to drink sucrose concentrations with one concentration present, volume consumed is inversely related to sucrose concentration. When a choice of concentration is permitted *S*, volume consumed is directly related to concentration. In that study, no change in amount consumed accompanied the transition from single-concentration tests to choice tests for the higher concentration. However, drinking of the lower concentration was drastically reduced in the choice tests.

On the assumption that lick rate is positively related to volume consumed, the predictions which follow from Premack's (1959, 1961) position

[1] This investigation was supported by Research Grants M-817 and M-3345 from the National Institute of Mental Health, of the National Institutes of Health, Bethesda, Maryland.

[2] Now at Florida State University.

are quite clear. First, when only one concentration is present, dependent rate will be inversely related to the concentration. Second, when a choice of concentration is available, dependent rate will be positively related to concentration. Finally, in the choice situation, dependent rate for the higher concentration will be equal to the dependent rate for this concentration when tested singly, whereas dependent rate for the lower concentration will be markedly reduced.

Method

Subjects.—The *S*s were four naive female albino rats approximately 110 days old at the start of the experiment.

Apparatus.—The apparatus used was a Skinner box with inside dimensions of $8\frac{3}{4} \times 6 \times 7\frac{1}{4}$ in. The four sides and top of the box were constructed of $\frac{1}{4}$-in. Masonite. The floor of the box was $\frac{1}{2}$-in. hardware cloth. Two levers, which could be presented singly or together, were mounted on one wall of the box 3 in. above the floor. The levers were $2\frac{1}{2}$ in. apart, separated by a $\frac{1}{4}$-in. piece of Plexiglas, and protruded $\frac{3}{4}$ in. into the box. Two liquid reinforcement magazines and two dippers delivered sucrose concentrations through openings in the floor of the box beneath the lever pressed. Each lever press activated only the dipper beneath the lever pressed.

Procedure.—One week prior to the start of the experiment proper, all *S*s were placed on a 23-hr. food deprivation schedule that was maintained throughout the experiment.

Two days of magazine training were administered. Each session was 40 min. long and consisted of 40 presentations of each dipper in random order. For each *S* one dipper presented 8% sucrose concentration and the other 32% throughout the experiment. Concentration was counterbalanced over *S*s with respect to position of the lever.

Twenty sessions of single-lever pressing were then given. Each session was 20 min. in duration and each press reinforced. An ABBA order was employed in determining which lever and associated concentration was presented during a session. The intersession interval was 24 hr. for the first nine sessions, 48 hr. for all remaining sessions.

Following single-lever sessions, seven 20-min. sessions with both levers and both associated concentrations were administered.

TABLE 1

Analysis of Variance of Mean Bar Presses per Session

Source	*df*	*MS*	*F*
Concentration (C)	1	3329.29	5.61*
Type of session (T)	1	18975.06	31.98**
C × T	1	22455.71	37.84**
Error	12	593.37	

* $P < .05$.
** $P < .001$.

Results and Discussion

The mean number of lever presses for each *S* on the last five single-lever sessions of each concentration and the mean number of presses on each lever for the seven double-lever sessions were determined. An analysis of variance was then carried out on these data. The summary of the analysis presented in Table 1 shows that the interaction between concentration and type of session is highly significant ($P < .001$). Figure 1 presents a plot of this interaction. As was predicted, the dependent rate for the lower concentration was greater than that for the higher concentration in single-lever tests. In double-lever tests, however, the rate for the lower concentration was drastically reduced

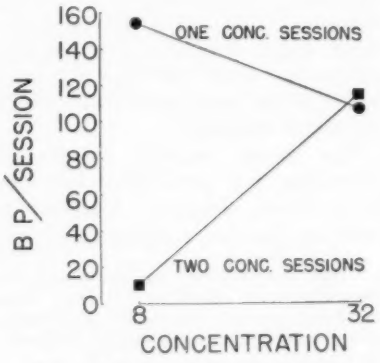

Fig. 1. Bar pressing as a function of concentration and type of session.

(approximately 1/15 of that obtained in single-lever tests). The rate on the higher concentration is approximately equal in both types of tests.

In support of the analysis, examination of individual data shows that for 3 of the 4 *S*s during the five single-lever sessions the lowest rate obtained on 8% contingent pressing was higher than the highest rate obtained on 32% contingent pressing. On each of the seven double-lever sessions the lowest rate obtained on the 32% contingent pressing was higher than the highest rate obtained on the 8% pressing for all *S*s.

The results of this study clearly support Premack's (1959, 1961) assumptions regarding the independent rate of the contingent response as a determiner of the dependent rate. An unequivocal relationship between independent rate of contingent response and dependent rate was demonstrated in that reversal of the independent rate of the contingent response was paralleled by the reversal in the dependent rate.

Recently, Premack (1961) has demonstrated a reversal in dependent rate accompanying a reversal in the independent rate of drinking on various sucrose concentrations where the parameter of concern was duration of licking prior to introduction into the contingency situation. As has been noted (Premack, 1961), demonstrations with the consummatory response contribute little to testing the assumption that a rate differential is a necessary condition for reinforcement. Nevertheless, support for the sufficiency of the condition accrues as the parameters governing the rate of the consummatory response function are shown to yield congruous consummatory and dependent rate functions.

Finally, returning to the parameter of concern in the present study, the traditional approach to this problem has been to ascertain the bar-press rates in the single-bar situation and assume some mode of interaction, e.g., ratio, subtractive, etc., to predict rate in the choice, or conflict, or competition, or selective learning situation. It is clear that no interaction rule as yet offered could predict the superiority of the 32% contingent bar. Since no change in the variables determining the prechoice response strengths, maintenance schedule, reinforcement magnitude, etc., accompanied the transition to the choice situation, the direction of the prechoice difference would be expected to be maintained.

Summary

Four *S*s were given bar-press tests with 8% and 32% sucrose concentrations presented singly and simultaneously. When tested singly, bar-press rate was inversely related to concentration. When tested simultaneously, a direct relationship was obtained. The direct relationship is due to the much suppressed responding on the 8% contingent bar with no change in rate of the 32% contingent bar. The fact that the bar-press rates parallel rates obtained for drinking in single and simultaneous tests supports Premack's (1959, 1961) theory.

References

Premack, D. Toward empirical behavior laws: I. Positive reinforcement. *Psychol. Rev.*, 1959, **66**, 219–233.

Premack, D. Predicting instrumental performance from the independent rate of the contingent response. *J. exp. Psychol.*, 1961, **61**, 163–171.

Young, P. T., & Greene, J. T. Quantity of food ingested as a measure of relative acceptability. *J. comp. physiol. Psychol.*, 1953, **46**, 288–294.

(Received October 5, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 32–35

MAGNITUDE OF REWARD AND PROBABILITY LEARNING [1]

YVONNE BRACKBILL, MICHAEL S. KAPPY, AND RAYMOND H. STARR

Johns Hopkins University

A first concern in studies of probability learning—and still a persistent one—is the "matching" choice behavior that *S*s show when confronted with events of unequal occurrence. Instead of behaving like rational game theorists by learning to respond at unity, *S*s tend to match the relative frequencies of their responses to the relative frequencies of the stimulus events (Grant, Hake, & Hornseth, 1951). Goodnow (1955) has advanced an interesting set of hypotheses concerning factors that lead to matching; she cites intolerance for boredom, inability to overlook losses, and so on. Statistical learning theories provide a more formal approach to the prediction of probability matching (cf. Bush & Estes, 1959).

In terms of experimental evidence, one variable that appears to be of no small importance in predicting and manipulating relative response frequencies is the nature of tangible reinforcing contingencies offered by the experimental procedure. First, the mere introduction of reward for correct predictions has been shown to increase maximum gain responding. Money may not be everything, but it can do a lot toward alleviating the boredom that comes from making the same verbal response for 100 trials or more (Siegel & Goldstein, 1959). Further, it has been found that maximum gain responding is increased by the use of both reward for correct prediction and punishment for incorrect prediction (as in the "risk" condition of Siegel & Goldstein, 1959), by a previous history of infrequent rewards (or low "expectancy for success," Stevenson & Zigler, 1958), and by higher frequencies of reward in a contingent procedure (Stevenson & Weir, 1959).

In the present study, the effect of amount of reward was investigated, with the expectation that *S*s will maximize gain to the extent that there is something tangible to be gained by so doing. In other words, the hypothesis was that in the typical noncontingent probability learning situation, maximum gain strategy is a direct function of the amount of reward offered for correct predictions.

METHOD

Subjects.—The *S*s were 24 boys and 24 girls from the second grade classes of a single elementary school. Mean age was 95.5 mo., $SD = 9.4$. Mean IQ was 85.8, $SD = 18.9$, by the SRA test. The sample was relatively homogeneous in terms of socioeconomic background; modal occupation level for fathers corresponded to Class 5 of Warner's (1949) 7-point scale. Thirty-four *S*s were white, 13 were Negro, and 1 was Oriental.

Experimental material.—This consisted of 3×5 in. cards cut from green, light weight cardboard stock. On each card was printed the picture of either a cat or a dog. Two packs of 200 cards each were prepared, in order that dog or cat would appear equally as the more frequent event over treatment levels, sex of *S*s, and sequence of running *S*s. In preparing these packs, the relative proportions of the two stimulus events were held

[1] This investigation was carried out during the term of a Postdoctoral Fellowship awarded to the first author by the National Institute of Mental Health. The authors are grateful for the cooperation of Mary Adams, Dorothy Rawlings, and the staff of the Barclay Elementary School, Baltimore, Maryland.

constant for all treatment levels at .75:.25. A randomized series of events was prepared (and recorded) in blocks of 20 trials each, with the overall objective probability split of .75:.25 maintained within each block.

There was no other restriction on randomization. Twenty such 20-trial blocks were prepared. Then, for each *S*, a 10-block sequence was randomly drawn from the 20-block pool.

In addition to the pack of 200 cards, a box containing approximately 750 marbles and a cup with a 100 marble capacity in which *S* accumulated his winnings were also on the table at which *E* and *S* worked. On display in another part of the room was a standard array of ten 10¢ toys. These were high-preference toys, selected on the basis of previously obtained rankings from same-age *S*s.

Procedure.—Forty-eight *S*s were randomly assigned to four experimental groups, with the single restriction that each group of 12 contain equal numbers of boys and girls. The four groups were given 5, 3, 1, or 0 units of reward (marbles) for each correct response. One hundred marbles could be exchanged for one toy.

At the beginning of a series, the pack of 200 cards was face down; *S* was instructed to guess before *E* turned each card over, whether the card would show a dog or a cat. Both at the beginning of the experiment and subsequently, whenever he had accumulated 100 marbles, *S* was asked if he wanted to try to win a toy (or another toy). If he responded negatively at any choice point, the procedure was discontinued. One prospective *S* was discarded at the beginning of the experiment after stating that none of the available toys interested her.

Results and Discussion

Figure 1 shows the mean proportions of predictions of the more frequent event ($S_{.75}$) in successive 20-trial blocks for the four groups. As a test of the hypothesis that maximum gain responding increases as a function of magnitude of reward for correct predictions, the Jonckheere (1954) test against ordered alternatives was applied to the data from the last three blocks. The resulting statistic of 2.65 is significant at the .004 level by a one-tailed test. (Scores from three blocks were used in order to minimize the number of tied scores. If one used the same test on scores of the last block only and split the ties, the resulting statistic of 1.83 would still be significant at the .02 level.)

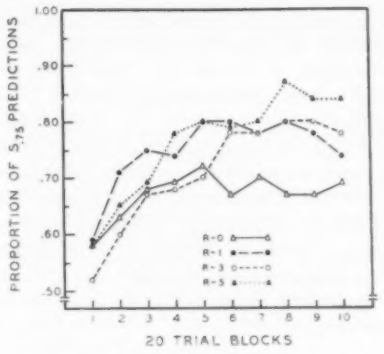

FIG. 1. Mean proportions of $S_{.75}$ predictions in successive 20-trial blocks for four experimental groups differing in number of units of reward offered for every correct prediction. (Each plotted point is a Group × Block mean based on 12 cases.)

Inspection of Fig. 1 shows that, for the last half of the trials, the most noticeable difference among treatment levels is between nonreward and any value of reward. This suggests that the performance function under amount of reward is not linear for these data. The same suggestion has been made by Pubols (1960), who concluded from his review of studies on amount of reward in learning and performance in animals, that the performance function ". . . is negatively accelerated, and possibly logarithmic . . ." (p. 111). If one prefers a technique more sophisticated than simple visual inspection, the same question may be considered from the standpoint of utility theory. On the basis of Siegel's (1959) Model I (Equation 1), it is possible to calculate the ratio of the utility of making correct responses to that of vary-

ing responses (assuming constant utility of varying responses). The resulting α values for R-0 through R-5 (and interpolating for R = 2 and R = 4) would be .72, 1.08, 1.13, 1.16, 1.29, and 1.40. Clearly, the largest increment in utility value follows the introduction of any material reward; subsequent additions in reward produce relatively less effect on utility value, so that the resulting utility curve becomes negatively accelerated.

Sequence Analysis

A first-order sequence analysis was performed on the last 100 trials for each *S*. Such an analysis yields the probability of *S*'s prediction on Trial *n*, given two items of information about Trial *n*-1: (*a*) the event that *S* had predicted previously, and (*b*) the event that had actually occurred previously. Since the Jonckheere (1954) analysis had already shown tangible rewards to have a significant effect on responding, it is reasonable to expect that previous actual occurrence rather than previous prediction would emerge from a sequence analysis as the more dependable predictor from Trial *n*-1 to *n*. This expectation finds support in a comparison of Columns A and B, Table 1; the average difference between the two sets of probabilities is a mere .04. Further, the ordering within the two sets of probabilities reflects with fair accuracy the effect of the experimental variable, magnitude of reward.

However, comparison of Columns C and D reveals a different outcome. Prediction to Trial *n* is not a simple function of previous actual occurrence on Trial *n*-1, as it appears to be for the A-B comparison. In fact, the mean difference between the two sets of 48 correlated proportions is significantly different from zero ($CR = 2.82$, $P < .01$). This means that, for those trials on which the less frequent event actually occurred, not only that occurrence but also *S*'s own previous prediction determined his next prediction. Specifically, *S* was much more willing to revert to a maximum gain strategy on his next prediction *if* he had just successfully predicted the less frequent event than if he had not predicted it and it had occurred. This effect appears to be independent of the experimental variable, since there is no systematic trend among C-D difference scores as a function of treatment level.

What is the interpretation of this effect, and how does it relate to the experimental variable? Obviously, the best

TABLE 1

PROBABILITY OF PREDICTING $S_{.75}$ ON TRIAL *n*, GIVEN CERTAIN INFORMATION ABOUT TRIAL $n - 1$

Group	Event on Trial $n - 1$			
	A *S* Predicted $S_{.75}$ & $S_{.75}$ Occurred	B *S* Predicted $S_{.75}$ & $S_{.75}$ Occurred	C *S* Predicted $S_{.25}$ & $S_{.25}$ Occurred	D *S* Predicted $S_{.75}$ & $S_{.25}$ Occurred
R-0	.74	.82	.56	.38
R-1	.80	.83	.85	.65
R-3	.82	.87	.73	.67
R-5	.89	.87	.78	.64

Note.—Each entry represents the mean of 12 individual proportions.

strategy to maximize tangible reward is for *S* to predict the more frequent event on every trial. Further, this strategy is the more compelling the more the reward. That *S*s do behave in accordance with this strategy was shown by the Jonckheere (1954) analysis. What the sequence analysis shows is that there is a second source of reinforcement at work: the utility to *S* of correctly predicting the occurrence of the less frequent event. Where the investigators had intended only one game, *S*s discovered two.

Whether or not one agrees with this interpretation, the fact remains that the net result of this secondary effect is to subtract from maximum gain responding. In effect, it constitutes a response that is competing with the criterion response. However, in contrast with the results of studies in which incentive magnitude has been found to affect both criterion response and competing responses (Pereboom & Crawford, 1958), the present results show only the criterion response to have been influenced by magnitude of reward. The competing response appears to have been maintained by some other form of reinforcement, independent of the form under study.

Summary

This study investigated probability learning in a binary choice, noncontingent procedure with 0, 1, 3, and 5 units of reward for each correct prediction, 100 units being exchangeable for one toy. The *S*s were second-grade children. Maximum gain responding increased with larger values of reward, although the function relating these two variables did not appear to be linear. From a first order sequence analysis, an independent effect emerged through an unexpected source of reinforcement: the utility to *S* of correctly predicting the occurrence of the less frequent event.

References

Bush, R. R., & Estes, W. K. (Eds.) *Studies in mathematical learning theory.* Stanford, Calif.: Stanford Univer. Press, 1959.

Goodnow, J. J. Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.

Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

Jonckheere, A. R. A distribution-free *k*-sample test against ordered alternatives. *Biometrika*, 1954, **41**, 133–145.

Pereboom, A. C., & Crawford, B. M. Instrumental and competing behavior as a function of trials and reward magnitude. *J. exp. Psychol.*, 1958, **56**, 82–85.

Pubols, B. H., Jr. Incentive magnitude, learning, and performance in animals. *Psychol. Bull.*, 1960, **57**, 89–115.

Siegel, S. Theoretical models of choice and strategy behavior: Stable state behavior in the two-choice uncertain outcome situation. *Psychometrika*, 1959, **24**, 303–316.

Siegel, S., & Goldstein, D. A. Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychol.*, 1959, **57**, 37–42.

Stevenson, H. W., & Weir, M. W. Variables affecting children's performance in a probability learning task. *J. exp. Psychol.*, 1959, **57**, 403–412.

Stevenson, H. W., & Zigler, E. F. Probability learning in children. *J. exp. Psychol.*, 1958, **56**, 185–192.

Warner, W. L. *Social class in America.* Chicago: Science Research Associates, 1949.

(Received October 10, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 36–41

THE INFLUENCE OF STRENGTH OF DRIVE ON FUNCTIONAL FIXEDNESS AND PERCEPTUAL RECOGNITION [1]

SAM GLUCKSBERG [2]

New York University

The present study represents an attempt to conceptualize the functional fixedness phenomenon (e.g., Adamson, 1952; Duncker, 1945) in terms of an S-R model. Historically, this phenomenon has been represented as reflecting perceptual processes. For example, when *S* uses a matchbox as a platform in order to solve the problem of mounting a candle on a vertical screen (Adamson, 1952), he is described as being able to perceive the box as a platform. Failure of problem solution is interpreted as indicating that *S* can only perceive the container properties of the box and not its platform properties.

In a certain sense (Lachman, 1960) the model used to represent behavior is arbitrary. The same functional fixedness problem can be conceptualized in terms of competition between S-R habits, using the box as a container and using it as a platform.

A number of studies conducted within the S-R orientation (Maltzman, Fox, & Morrisett, 1953; Spence, Farber, & McFann, 1956; Spence, Taylor, & Ketchel, 1956) have shown that a high drive can impair performance when the correct response is low in the habit hierarchy. One reason for this is that high drive prolongs extinction of the dominant habit (Perin, 1942) and thus retards the correct habit from gaining ascendancy (Kendler & Lachman, 1958). According to this analysis, if it were possible to control the position of the correct habit (using the box as a platform) in the response hierarchy, then it would be possible to control the course of problem solution. When the correct habit is low in the habit hierarchy, a high drive should retard problem solving. In contrast, when the correct habit is dominant, a high drive should facilitate problem solving. Experiment I tests this analysis. Experiment II extends the analysis to perceptual recognition.

EXPERIMENT I

Method

Subjects.—The *S*s were 139 male undergraduates randomly selected from the introductory psychology course at New York University. Eleven *S*s were discarded; 5 reported familiarity with the experimental task, and 6 had weak color vision. These latter 6 *S*s could not discriminate between red and black at an exposure duration of 20 msec., and could not be used in the perceptual recognition study which is reported as Exp. II. The total sample thus consisted of 128 *S*s.

These 128 *S*s served in both experiments. Half the *S*s were initially given the perceptual recognition tests, while the remaining *S*s had the problem solving task administered first. Three to 4 weeks intervened between successive testing sessions.

Experimental design.—A simple 2×2 factorial design was employed involving two levels of drive and two different response hierarchies in a Duncker-type (1945) candle problem. In this problem *S*s are instructed

[1] This paper is based on a PhD dissertation submitted to the Graduate School of Arts and Sciences, New York University. The author is greatly indebted to H. H. Kendler for his guidance and encouragement during the course of the study, and to the Office of Naval Research for support Contract Nonr 0285(30), H. H. Kendler, principal investigator.

[2] Now at United States Army Ordnance Human Engineering Laboratories, Aberdeen Proving Ground, Maryland.

to mount a candle on a wall. Available to them are cardboard match and thumbtack boxes, matches, and thumbtacks. The problem can be solved by tacking an emptied box to the wall and placing the candle in or on it.

Drive was manipulated by offering different incentives and the response hierarchies were controlled by the manner in which the problem was arranged. When the boxes are presented empty, the correct response is initially high in the hierarchy (hereafter referred to as Cond. Dm+), and the problem is easily solved (Adamson, 1952; Duncker, 1945). When the boxes shown to *Ss* are filled with tacks and matches, the correct response is initially low in the hierarchy (Cond. Dm−), and relatively few *Ss* succeed in solving the problem.

Experimental materials.—The following materials were presented on a table in standard positions in front of *S*: a round thumbtack box, approximately 50 thumbtacks, a rectangular pocket-size matchbox, approximately 35 safety matches, and one wax candle. A large box was used to conceal the materials while *S* received instructions. A corrugated paper board, 36 × 24 in., was mounted on the plaster wall of the laboratory. This board was removable, and served as the "wall" in the experiment. Each *S* was provided with fresh, unmarked materials.

Procedure.—The *Ss* were seated at the table and given instructions on the requirements of the problem. In addition, low-drive *Ss* were told: "We are doing pilot work on various problems in order to decide which will be the best ones to use in an experiment we plan to do later. We would like to obtain norms on the time needed to solve."

In place of the above instructions given to low-drive *Ss*, high-drive *Ss* were told: "Depending on how quickly you solve the problem you can win $5.00 or $20.00. The top 25% of the *Ss* in your group will win $5.00 each; the best will receive $20.00. Time to solve will be the criterion used."

When *S* had indicated that he understood the instructions the covering box was removed, *E* said, "begin" and a stop watch was started. Timing was stopped when *S* had either pushed a tack through an empty box or had failed to solve within 15 min. A score of 15 min. was assigned to failure *Ss*.

Results and Discussion

The response measure used to evaluate the effects of drive level and response hierarchies on problem solving was time to solve, in minutes. The mean solution times, together with number of failures per group, are presented in Table 1. The results provide evidence consistent with the experimental predictions. When the correct response is dominant (Dm+) high drive tends to facilitate problem solving; when an incorrect response is dominant (Dm−) high drive tends to retard problem solving.

Even though solution time scores were skewed, analysis of variance and *t* tests were applied to the solution time scores (Boneau, 1960). The significant difference ($F = 28.42$, $df = 1/124$, $P < .001$) between Groups Dm+ and Dm− validates the experimental manipulation controlling the position of the correct and incorrect responses in the hierarchy. The Dm+ form of the problem, as expected, is easier than the Dm− form. The difference between the means of the over-all high- and low-drive groups was not significant. On the other hand, the predicted interaction between drive levels and nature of the dominant response was obtained ($F = 7.37$, $df = 1/124$, $P < .01$).

Differences between the means of high- and low-drive groups within response hierarchy conditions were

TABLE 1

SOLUTION TIME SCORES TO SOLVE CANDLE PROBLEM: EXP. I

Group		Number of Failures	Solution Time (Min.)	
Nature of Dominant Response	Drive Level		Mean	*SD*
Dm+	High	2	3.67	4.52
	Low	6	4.99	5.42
Dm−	High	16	11.08	4.84
	Low	10	7.41	5.63

evaluated by t tests. The difference was significant ($t = 2.73$, $P < .01$) in Cond. Dm− but not in Cond. Dm+.

The failure to obtain a significant effect in Cond. Dm+ may be attributed, in part, to the fact that the performance of the correct response may have been too easy, producing a ceiling effect. Drive, as an energizer, is expected to increase speed of responding. The correct response, when it occurred, was executed fairly rapidly, usually within 5–10 sec. The effect of drive upon solution time was thus limited by the simplicity of the behavior required. Also, the response hierarchies for all *Ss* may not have been adequately controlled. The facilitative effect of drive could well have been masked because the correct response was not dominant for some *Ss*. For some *Ss* this was obviously the case; 8 of the 64 *Ss* in Group Dm+ failed to solve (see Table 1). Some *Ss* who did solve the problem attempted initially to melt wax and use the wax as an adhesive. This response, of course, is incorrect. If the dominant response is not the correct response, as the above suggests, then drive should not facilitate problem solving. Thus the lack of significance might be due to the fact that the results obtained from *Ss* whose dominant response was truly the correct response were offset by the results of *Ss* whose initial responses were incorrect. It would seem, if one is to judge by the above, to be extremely difficult to arrange a problem situation in which all *Ss* respond initially by making a correct response.

Experiment II

A representation of functional fixedness within S-R drive theory (Spence, 1956) proved to be fairly accurate. The problem that is now posed is whether the same representation can be applied to a simple perceptual task. If the S-R model applies to problem solving behavior that has been traditionally represented in perceptual terms, then can the same model also represent a perceptual task itself? A perceptual task (word recognition) was designed to be analogous to the problem solving task. The effects of drive were investigated under two conditions: dominant response correct (Dm+), and incorrect (Dm+).

TABLE 2

Stimulus Words Used in Word Recognition Task: Exp. II

Dm+		Dm−	
PAINTER	CHURCH	ELEGIES	VIGNETTE
SAVINGS	SERVICE	FRUGALITY	AMICABLE
LAWYER	HEAVEN	ASSIDUOUS	CHANCELS
POETRY	MARKET	PERCIPIENCE	MUNDANE
EDUCATION	DEBATING	ERUDITION	UNCOERCED
PHYSICS	SYMPATHY	BEATIFIC	BENIGN

Method

Subjects.—The 128 *Ss* of Exp. I served in this study.

Experimental design.—The same 2×2 factorial design in Exp. I was used. A word recognition test was administered under one of two drive conditions (high or low). The task required either the recognition of an obvious word or one that was difficult to identify. The *Ss* confronted with the former task were members of the Dm+ groups. The latter belonged to the Dm− groups.

The words to be recognized were presented tachistoscopically. Two words were presented simultaneously on each trial. One word, high on the Thorndike-Lorge (1944) word count, was printed in black upper case letters, and was easy to recognize. The other was low on the Thorndike-Lorge word count, printed in pale red lower case letters, and was difficult to identify (Howes & Solomon, 1951).

Twelve such word pairs were presented. The words used are contained in Table 2. The *Ss* in the Dm+ groups were instructed to identify the black words while *Ss* in the Dm− groups were required to identify the red words. The *Ss* were instructed as follows:

> This is a tachistoscope. If you look into the eyepiece you will see a black cross on a white ground. That is your fixation point; any stimuli that appear will appear at that distance from your eyes in just about that general area. I am going to flash words on the screen. Each stimulus card will

have two words printed on it. One of the words will be printed in black, one in red. I am gradually going to increase the exposure time until you can tell me what the red (or black) word is. Remember, we are only interested in the red (black) word. Try to report the word as soon as you can. Wrong guesses will not be counted.

The *Ss* were assigned to one of the four experimental groups in the following manner. The 64 *Ss* whose initial experimental task was the perceptual recognition problem were randomly distributed equally among the four groups. Subsequently, the 16 *Ss* in each of these four groups were proportionally assigned, at random, to the four problem solving groups. The 64 *Ss* who were initially given the candle problem were assigned to the experimental groups of this study in a similar fashion. This procedure minimized systematic sequence effects and permitted separate statistical treatment of the two experiments.

Apparatus.—The 12 word pairs were presented in a modified Dodge tachistoscope, described in detail by Karlin (1955). This instrument permits exposure durations ranging from .001 sec. to 2 min.

Procedure.—The *Ss* received instructions as described above, and were then differentiated into high- and low-drive groups by instructions similar to those given in Exp. I, i.e., high-drive *Ss* were offered money, low-drive *Ss* were not.

When the instructions had been given, visual duration thresholds were taken for each of the 12 words by the ascending method of limits, in 5-msec. steps. A current of 20 ma. was supplied to the field and target lamps.

The procedure followed was that described in detail by Howes and Solomon (1951) with the following exceptions: (*a*) the criterion for correct recognition was one correct verbal report instead of three consecutive reports; (*b*) since the thresholds for the individual words were of no interest the word order for all *Ss* was the same; and (*c*) no practice trials were given. The score for each *S* was the mean visual duration threshold for all 12 words.

Results and Discussion

The mean and median thresholds (in msec.) are presented in Table 3. All differences obtained between groups were in the predicted directions. In Cond. Dm+ high-drive *Ss* were superior to low-drive *Ss*. The direction of the difference was reversed in Cond. Dm−. Because of the marked heterogeneity in the group variances, an analysis of variance was made on log transformed scores. The means and *SD*s of the transformed scores, together with the relevant *t* tests, are presented in Table 4.

TABLE 3

MEAN AND MEDIAN VISUAL DURATION THRESHOLDS (MSEC.)

Group		Visual Duration Thresholds			
Nature of Dominant Response	Drive Level	Mean	*SD*	Median	Range
Dm+	High	34.0	14.2	29.5	17.0– 75.0
	Low	47.0	27.1	41.0	14.0–145.0
Dm−	High	199.8	69.8	207.5	77.0–347.0
	Low	151.9	83.7	133.0	42.0–430.0

The significant difference between Groups Dm+ and Dm− ($F = 330.70$, $df = 1/124$, $P < .001$) serves to validate the experimental manipulation of response hierarchies, i.e., Cond. Dm+ is easier than Cond. Dm−.

As expected, the difference between the over-all means of the high- and low-drive groups was not significant.

TABLE 4

MEANS, VARIANCES, AND *t* TESTS OF LOG_{10} VISUAL DURATION THRESHOLDS

Group		Log Visual Duration Thresholds		
Nature of Dominant Response	Drive Level	Mean	S^2	*t*
Dm+	High	0.4909	0.8658	2.46*
	Low	0.6154	1.6407	
Dm−	High	1.2708	0.9003	2.89**
	Low	1.1234	1.6054	

[a] Mean square of within group variance used as estimate of population variance.

* $P < .02$.

** $P < .01$.

The predicted interaction between drive level and nature of the dominant response was significant ($F = 14.30$, $df = 1/124$, $P < .001$). The differences between the means of transformed scores of high- and low-drive groups within response hierarchy conditions were evaluated by t tests (see Table 4). The high-drive *Ss* have significantly lower thresholds under Cond. Dm+ ($P < .02$) and significantly higher thresholds under Cond. Dm− ($P < .01$).

The question arises, what would have happened if, in Cond. Dm−, the red word had been presented alone? The recognition of the black word, including attending to it, was assumed to be the incorrect dominant response. However, perceiving the black word is not the only incorrect response possible for Group Dm−. The initial response to the red word can also be incorrect, and when it is it must undergo extinction. The *Ss* in the Dm− groups consistently misreported the red words: BEATIFIC was given as BEAUTIFIC or BEAUTIFY, ELEGIES as ELEGANT, CHANCELS as CHANNELS, etc. Furthermore, the errors tended to persist over several trials, in one case for 20 trials until the word (CHANCELS) was correctly identified.

The presence of the black word, then, was probably unnecessary. That is, the same results would have been obtained for Group Dm− if only the red words had been exposed. Since the dominant response (the misreport of the word) would still be incorrect, a high drive should prolong its extinction and thus impair correct recognition. Following this line of argument, it would be predicted that there is an interaction between drive level and word frequency. If an inverse relationship is postulated between word frequency and probability of misreport (i.e., as word frequency decreases the probability of erroneous report increases), then a high drive will impair recognition of low frequency words. Conversely, a high drive will facilitate recognition of high frequency words when the probability of error is low.

In conclusion it should be noted, first, that neobehavioristic drive theory has been successfully extended to two different areas. In each case the major prediction of the theory was confirmed: there is an interaction between drive level and nature of the response hierarchy. Secondly, the extension of the theory could be made only by representing functional fixedness and perceptual recognition in S-R terms. Once this had been done a limited number of principles derived from conditioning and simple selective learning sufficed to account for the obtained effects of drive. Of course, the phenomena concerned are not to be equated with conditioning, even though the conditioning model proved useful (Lachman, 1960). Instead, it is suggested that important empirical consequences have been obtained, indicating the utility of an S-R approach to problem solving.

Summary

A problem solving task and a perceptual recognition task were represented in S-R terms. Predictions derived from neobehavioristic drive theory were then tested in each situation. Each task was presented under one of two drive levels (high or low) and in one of two forms (dominant response correct or incorrect). High drive impaired performance in both tasks when the dominant response was incorrect. When the dominant response was correct high drive facilitated performance in the perceptual recognition task. The results in the problem solving task, although in the predicted direction, failed to achieve statistical significance. The major prediction of the theory, that there is an interaction between drive level and nature of the dominant response, was confirmed.

References

ADAMSON, R. E. Functional fixedness as related to problem solving: A repetition of three experiments. *J. exp. Psychol.*, 1952, **44**, 288–291.

BONEAU, C. A. The effects of violations of assumptions underlying the t test. *Psychol. Bull.*, 1960, **57**, 49–64.

DUNCKER, K. On problem-solving. *Psychol. Monogr.*, 1945, **58,** (5, Whole No. 270).

HOWES, D. H., & SOLOMON, R. L. Visual duration thresholds as a function of word-probability. *J. exp. Psychol.*, 1951, **41,** 401–410.

KARLIN, L. The New York University tachistoscope. *Amer. J. Psychol.*, 1955, **68,** 462–466.

KENDLER, H. H., & LACHMAN, R. Habit reversal as a function of schedule of reinforcement and drive strength. *J. exp. Psychol.*, 1958, **55,** 584–591.

LACHMAN, R. The model in theory construction. *Psychol. Rev.*, 1960, **67,** 113–129.

MALTZMAN, I., FOX, J., & MORRISETT, L., JR. Some effects of manifest anxiety on mental set. *J. exp. Psychol.*, 1953, **46,** 50–59.

PERIN, C. T. Behavior potentiality as a joint function of the amount of training and degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30,** 93–113.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W., FARBER, I. E., & MCFANN, H. H. The relation of anxiety (drive) level to performance in competitional and non-competitional paired-associates learning. *J. exp. Psychol.*, 1956, **52,** 296–305.

SPENCE, K. W., TAYLOR, J., & KETCHEL, R. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp. Psychol.*, 1956, **52,** 306–310.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

(Received October 11, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 42–49

PREFERENCE PROBABILITY BETWEEN GAMBLES AS A STEP FUNCTION OF EVENT PROBABILITY[1]

R. DUNCAN LUCE AND ELIZABETH F. SHIPLEY

University of Pennsylvania

This experiment is another of the studies arising ultimately from the utility ideas of Ramsey (1931) and von Neumann and Morgenstern (1947). It is based on a probabilistic analysis of choices among uncertain outcomes that rests upon two major assumptions. (Luce, 1959, Ch. 3). One, a general choice axiom, relates the choice probabilities for any set of three alternatives to those for its two-element subsets. Specifically, it postulates that if one of the three alternatives is eliminated, its probability is distributed between the other two in proportion to their probabilities in the original set of three. The second, a decomposition axiom (Luce, 1958; Luce & Raiffa, 1957) which has meaning only for uncertain outcomes, assumes that when S is choosing between certain pairs of gambles his preferences between the pure outcomes are statistically independent of his likelihood judgments between the chance events. More exactly, it says that if one knows the objective probability that one pure outcome is preferred to another and the objective probability that one event is deemed more likely to occur than another, then the probability of choosing between the gambles generated by these outcomes and events is given by the straightforward probability calculations under the assumption of independence. Although nothing is assumed in this postulate about the existence of subjective probability or utility functions as in classical utility theory, it is of the same spirit as the expected utility hypothesis, but much weaker. The precise mathematical statements of these two assumptions can be found in Luce (1959); they will not be repeated here.

Some psychophysical evidence supports the choice axiom (Anderson, 1959; Clarke, 1957); however, as yet, no direct preference tests have been made. Coombs (1958, 1959) presents indirect negative evidence, but there is some doubt about its meaning (Luce, 1959, p. 19). No data have yet been published on the decomposition axiom. On both counts, therefore, data are needed. Because these assumptions are difficult to test separately, and because together they imply a surprising result, we shall test only their conjunction.

Consider the one-person game:

	Option I	Option II
ρ	a	c
$\bar{\rho}$	b	d

where a, b, c, and d are sums (possibly negative) of money such that $a > c > d > b$ and where ρ and $\bar{\rho}$ mean the occurrence and nonoccurrence, respectively, of a certain chance event. The S selects one of the two options and then, following his choice, the chance event is given an opportunity to occur. These two decisions—one by S, one by chance—determine a cell in the matrix, which contains S's payoff. Thus, if S chooses Option II and the chance event fails to occur (row $\bar{\rho}$), the payoff to S is d.

[1] This research was supported in part by Grants NSF G-4506 and G-5544 from the National Science Foundation to Harvard University. The experimental work was carried out at the Laboratory of Social Relations, Harvard University.

Now, holding the payoffs fixed, consider S's choices as we vary the event probability. If the event is certain to occur—or so S believes—then it is only sensible for him always to choose Option I because chance will choose Row 1 and, by assumption, S prefers a to c. On the other hand, if the event is certain not to occur, S surely will never choose Option I because $b < d$. And as the event probability is varied from zero to one, S's probability of choosing Option I must change from zero to one. The question is: what is the nature of this transition?

Classical economic thought has assumed a simple discontinuity, but most psychologists doubt that this is correct, and the evidence (Chipman, 1960; Coombs, 1958, 1959; Davidson & Marschak, 1959; Davidson, Suppes, & Siegel, 1957; Mosteller & Nogee, 1951; Papandreou, 1957), although open to other interpretations, is generally viewed as being negative. Assuming that the function is not a simple discontinuity, the most plausible a priori conjecture is a smooth ogival function, so familiar from psychophysics. Plausible, yes, but not possible for all choices if the two assumptions mentioned above are correct. For payoff matrices where one row dominates the other (e.g., $a > b$ and $c > d$), they imply that the choice probability varies from zero to one by a series of jumps. That is to say, the function must be a step function, of which the simple discontinuity is a special case.[2] When neither row dominates the other, i.e., when the two largest payoffs lie on a diagonal, no theoretical prediction has been derived. It is not known whether steps are again necessary or whether continuous functions are possible.

Although this prediction of steps is qualitatively startling and sharply at variance with most psychologists' intuitions, it is uncertain whether it can be successfully tested—not only would negative results be completely ambiguous, but seemingly positive ones are open to other interpretations. Basically, there are three difficulties. First, the theory states neither how many steps to expect nor where to find them. If there are enough, say 8 or 10, reasonably spaced over a sufficiently narrow region of event probabilities, then a vast amount of data is needed to tell a step function from a smooth ogive. Second, to estimate the choice probabilities, the choices must be presented repeatedly, allowing all the well-known dangers of systematic changes over time. And third, even if by good luck and design these troubles are avoided, one must still live with the inevitable and considerable binomial variance of our observations which, on the one hand, may tend to mask the steps of a step function (unless they are broad) and, on the other hand, may introduce at least small steps into data arising from an ogive.

Severe though these difficulties seem to be, it seemed worthwhile to conduct the obvious experiment. The payoff, should the data hint at positive results, seemed sufficiently large to offset both the slim chance that the steps are there and the inherent uncertainties of knowing it even if they are.

Method

Subjects.—Without specifying the exact nature of the work, five Harvard College students were solicited to participate in an experiment for which they would be paid both an hourly wage and a bonus. After he had indicated interest in the work, each was asked if he had qualms about gambling; none did.

Apparatus.—The 2×2 payoff matrices were printed on small white cards. Each column defined a gamble, and so an option

[2] It should be noted that this conclusion, which is "global" in nature, follows from repeated application of Luce's (1959, p. 86) Theorem 13, which is "local" in nature. The theorem requires that various probabilities of choice be different from 0 and 1, which is what gives it its local character, but this assumption need not hold for any two arbitrarily chosen gambles of an experiment designed to test the step-function hypothesis. Thus, there can be no question of looking to the data to see whether prerequisites of the step-function hypothesis are satisfied.

for S; each row corresponded to one of two complementary events. Packs of these cards, face down, were mounted in a rack before S. Turning a card over to expose the matrix tripped a microswitch which in turn activated a pen of an Esterline Angus recorder. Two buttons, labeled to correspond to S's two options, were located in front of each S and were wired to the pen recorder.

The events were specified in terms of the faces shown by five dice. The possible outcomes were ranked into hands, much as in poker except that flushes could not occur, there being no suits, and straights were defined to be of the lowest rank. A hand known as the house's hand was printed on each card. One event was defined to have occurred when the dice exhibited a hand, known as S's hand, higher than the house's; the complementary event occurred when S's hand was the same as or less than the house's. Each S was provided with an ordered list of the 252 hands that might arise and of the probabilities of beating each of them. The events were produced by spinning a wire cage that contained the five dice.

Procedure.—Prior to the experiment, E explained both the procedure and the bonus system. The payoffs in the matrices were treated as points to be won or lost on each throw of the dice. At the end of the experiment the accumulated points would then be used to determine each S's share of a $200 bonus. They were told that a point was roughly equal to 1 cent, which was true.

An experimental trial was as follows. When E gave a ready signal to begin the trial, each S turned over the next card in his pile, examined the alternatives, and indicated his choice by pressing one of his two buttons. The time when the card was turned over, his choice, and the time it was made were all recorded. After all Ss had made their choices, one of them—the same one throughout an experimental session—spun the cage and called out the resulting hand. Each S then recorded on a sheet provided the points won or lost on the trial. They knew that on any given trial the payoff matrix and the house's hand differed among them, but that throughout the experiment they would all be presented with the same six matrices. They were also led to believe although this was not true, that they would all see the same house hands.

Four preliminary sessions were run to determine for each S-matrix combination the range of events in which S's choices were uncertain. The experiment itself consisted of daily sessions of 150 trials, lasting about 1 hr., for 33 days. Because some of the initial choices of event ranges for S_1 were inappropriate, he was run three extra sessions. For the first few weeks, sessions occurred only when all 5 Ss could attend; however, S_1 became ill, which made it necessary to relax this requirement. It seemed that the larger the group, the more enjoyable the session; some sessions at the end with a single S were considered dull. On the whole, however, if noisy rivalry is any sign, the Ss were interested and involved.

TABLE 1

PAYOFF MATRICES STUDIED

1		2		3	
I	II	I	II	I	II
+31	+22	+14	+5	− 4	−13
+10	+16	− 7	−1	−25	−19
4		**5**		**6**	
I	II	I	II	I	II
+33	+17	+11	−5	− 5	−21
+12	+23	−10	+1	−26	−15

Conditions.—The six payoff matrices used are shown in Table 1. The rational decision rule is the same for all: choose Option I if the probability of the first row event exceeds .4, and choose Option II if this probability is less than .4. The first three matrices differ from each other only by a positive or negative constant added to all entries. The entries are all positive in the first matrix, mixed in the second, and negative in the third. The last three matrices are similarly related. The first three meet the necessary inequalities so that a step function is predicted in response to them. For the second set of three, no prediction from the theory is known. They were run to see whether Ss appear to react differently to the two classes.

During the 4-day preliminary phase, seven events with probabilities from .2 to .8, spaced in .1 steps, were used. Each of the 42 event-matrix combinations was presented from 9 to 13 times. These data were used to estimate the approximate event probability for which S's choice probability was .5 for each matrix. A range from .1 below to .1 above this point was explored in the experiment. As mentioned above, some mistakes about S_1's ranges were made, but they were rectified shortly after the experiment began. Aside from that, all save one of the other ranges were appropriate.

Fifteen hands were selected in these .2 probability ranges so that the steps between adjacent ones varied from .020 probability at the edges to .008 in the center of the

range. Because we did not want *Ss* to be aware of differential treatment, events spaced at .05 intervals from .25 to .75 were also presented to all *Ss*. Nonetheless, S_4 and S_5 realized that the other *Ss* were getting different house hands, and they questioned *E* about this. They were assured that we would distribute the bonus on the basis of the number of points earned relative to the maximum possible.

Each of the 90 event-matrix combinations appeared once every 90 trials for a total of 50 presentations during the experiment. Within a block of 90, their order was randomized. The 36 event-matrix combinations that were added to this basic design to convince *Ss* that they were being treated alike were each presented five times. Rows or columns were interchanged randomly on successive presentations of the same event-matrix pair.

Results and Discussion

Plots of the estimated choice probabilities vs. the theoretical event probabilities for the six payoff matrices of the S_1, S_2, and S_3 are shown in Fig. 1. To save space, the corresponding plots for S_4 and S_5 are not presented because in all cases they are almost perfect simple discontinuities. For S_4 the break occurs at event $p = .4$, which is the rational decision, and for S_5 it occurs at $p = .5$. Thus, the data from S_4 and S_5 are consistent both with the step function hypothesis and with the classical economic assumption of a simple

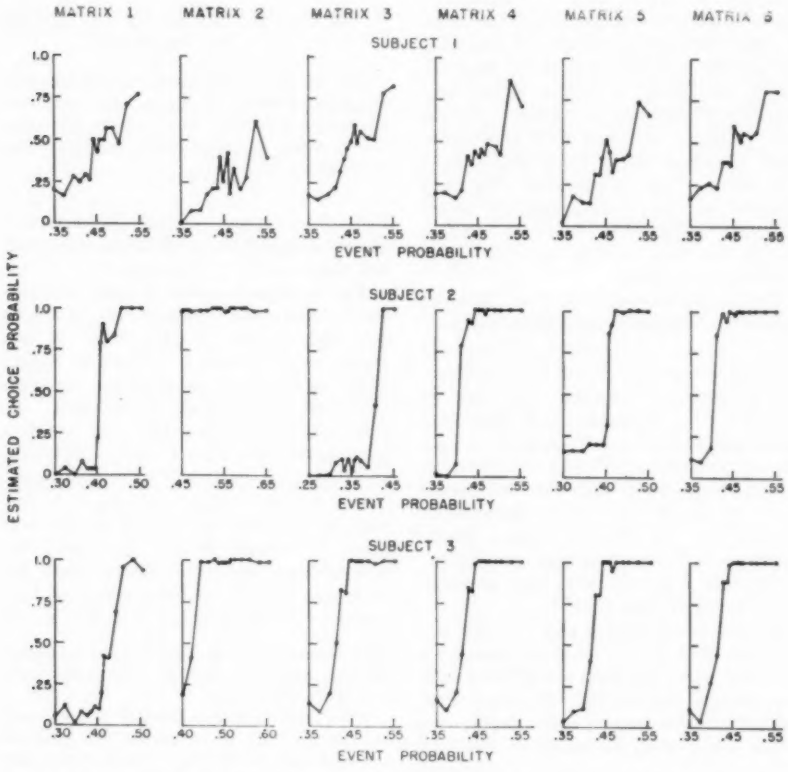

Fig. 1. Estimated choice probabilities vs. event probabilities for S_1, S_2, and S_3 on the six different payoff matrices. (Each point is based upon 50 observations.)

discontinuity, but they are not consistent with the ogival hypothesis (except, of course, in its limiting form). The data for *S*s 1–3 are certainly inconsistent with the assumption of a simple discontinuity, but whether steps or ogives underlie these data is less clear.

Because we do not know how to reduce this question to a well formulated and solved statistical problem, the reader ultimately will have to make his own decision. The following considerations may be helpful.

It is clear that the data exhibit plateaus, some of which are broad (particularly in S_1, Matrices 1–6; S_2, Matrices 3 and 5; and S_3, Matrix 1). There is also no question that the edges of these flats are rounded, but considering all the possible sources of blurring, this is the least to expect if the step function hypothesis is correct. So the only problem we need face is whether these plateaus could have arisen from a continuous, ogival function. We have thought of only two possible ways: binomial variability and an abrupt change in the location of the ogive. Both are discussed.

Three arguments can be brought against the binomial variability explanation. The first is simply the subjective impression that each *S* exhibits a characteristic style of behavior over the six matrices. For example, in S_3 there is a small jog at about $p = .425$ in five of the six matrices. The style seems to persist even when different ranges of event probabilities were employed (see S_2 and S_3). If the flat regions really result only from sampling variability, no internal *S* consistency, such as there seems to be, should exist.

Second, if variability accounts for the plateaus, then they should not stand up very well when the data are divided into two parts. Separating the even and odd observations for each point, we obtain plots such as those in Fig. 2.[3]

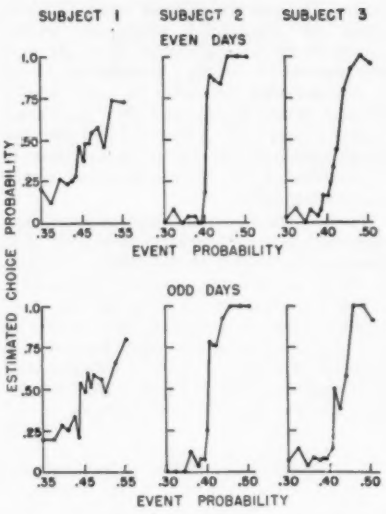

FIG. 2. Estimated choice probabilities vs. event probabilities for S_1, S_2, and S_3 on Payoff Matrix 1 for the even and the odd days. (Each point is based upon approximately 25 observations.)

Note that at least one of the narrower plateaus vanishes, as might be expected, but the broader ones remain.

Third, in the absence of more sophisticated statistical procedures, we can generate Monte Carlo data to see whether, by chance, plausible ogives could be expected to generate reversals of the magnitude seen in the data. To this end, we had the following problem programed for the Univac computer at the University of Pennsylvania.[4] Given $0 < p_1 < p_2 < \cdots < p_{15} < 1$, use random numbers generated in the machine to obtain N (in our case, 50) independent binomial observations for each p_i. Let N_i be the number of successes. For each lag k, $k = 1, 2, \ldots, 8$,

[3] As it is impractical to present all 36 of these curves, and as there seems to be no particularly rational way to choose a few from among them, we simply have presented the first matrix for each *S*. The same choice is made in Fig. 3.

[4] We wish to thank Donald Norman and A. Martin Wall for preparing the program and carrying out the runs and we would also like to express our appreciation to the Computer Center of the University of Pennsylvania for its cooperation.

TABLE 2

PROPERTIES OF THE LOGISTIC FUNCTIONS $P=1/1+e^{\sigma s}$) USED TO GENERATE THE MONTE CARLO DATA

Ogive Number	σ for which $P = 1/2$	jnd	σ Range
1	.45	.06	.35–.55
2	.40	.02	.30–.50
3	.40	.02	.35–.55

store the number of differences $N_{i+k}-N_i$ which are zero or negative. Repeat this M times (we used $M = 200$) and read out, for each k, how many cases exhibited 0, 1, . . ., and 15-k reversals, i.e., zero and negative differences.

Three sets of probabilities p_i were selected as follows. Three logistic curves, $1/(1 + e^{\sigma s})$, were chosen to approximate most of the data curves; they have the features shown in Table 2. By letting σ run over the event probabilities used in the experiment, a set of 15 probabilities was generated from each logistic.

For each S and each matrix, we determined for each lag k the number of times the choice proportion was the same or smaller for an event k steps above another event. Then, using the Monte Carlo data for the best fitting of the three ogives, we found the percentage of times that many or more reversals of lag k occurred in the Monte Carlo runs. Thus, for example, S_2 in Matrix 4 exhibited six reversals of lag 2, and the Monte Carlo runs for Ogive 3 showed five cases out of 200 with six or more reversals at lag 2. So the entry for this case is $5.100/200 = 2.5\%$. These estimated probabilities, which are summarized in Table 3, measure how likely it is that reversal structure of our data could have arisen by chance from the ogive. It should be kept in mind that the eight entries per row are not independent.

The patterns for S_2 and S_3 seem unambiguous; it is most unlikely that our data could have arisen from a logistic with a .02 jnd. The pattern is less clear for S_1, but we note that except for Matrix 3 there is always at least one lag for which the observed number of reversals has 5 chances or less in 100 of occurring if the ogive is the true function.

On the basis of these considerations, we doubt that the observed plateaus can have resulted simply from ogives plus sampling variability.

A second way a continuous ogive might give rise to a plateau is for the center of the ogive to be located at one event probability on some trials and at another on the remainder. If the absolute value

TABLE 3

PERCENTAGE OF 200 MONTE CARLO RUNS IN WHICH THE NUMBER OF REVERSALS OF LAG k EQUALED OR EXCEEDED THE NUMBER OBSERVED

S	Matrix	Ogive	1	2	3	4	5	6	7	8
1	1	1	2	97.5	3.5	34.5	*	2	*	*
	2	1	77.5	40	3.5	0	0.5	0	*	0
	3	1	77.5	40	28	34.5	*	*	*	*
	4	1	5	76.5	28	34.5	18.5	2	*	0
	5	1	17.5	40	3.5	6	0.5	*	*	*
	6	1	5	14	28	34.5	18.5	*	*	*
2	1	2	0	0	0	0	0	*	*	*
	4	3	0	2.5	1	0.5	1.5	0	0	2
	5	2	0	0	0	0	*	*	*	*
	6	3	0	2.5	0.5	0.5	1.5	2.5	8	2
3	1	2	14.5	0	0.5	0	0	0	0	*
	3	3	0	2.5	0.5	0.5	0	0	0	2
	4	3	0	2.5	0.5	0.5	1.5	2.5	8	*
	5	3	1.5	2.5	0.5	0.5	1.5	0	0	2
	6	3	0	2.5	0.5	0.5	1.5	2.5	8	*

Note.—The asterisk indicates that in these cases no reversals were observed, hence that number was equaled or exceeded in all Monte Carlo runs.

of the difference in locations is large relative to the breadth of the ogive, then the expected data would exhibit a single intermediate step with rounded edges. With three discrete locations for the ogive, two rounded steps can result. And so on. Such discontinuous shifts of location might, for example, correspond to major shifts in S's opinion about the rational breaking point. If these abrupt changes in location occurred irregularly from day to day during the experiment, we know no way of detecting them; however, if the shifts occurred only once or twice during the experiment, then, for example, the data from the first half of the experiment should differ appreciably from those from the second half. Such plots are shown in Fig. 3; the horizontal features of the plots are substantially the same in the two halves.

Although there is no evidence for the sort of shift that would produce plateaus, this figure does suggest that there may have been systematic changes in the underlying probabilities during the experiment. To check this, the choice proportions for the first and second halves of the data were compared for each S-matrix combination by the Wilcoxon matched-pairs signed rank test. The P levels are summarized in Table 4, where we see that in over half the cases there are significant trends. The effects are not, however, very large. Their main consequence would be to increase the variance of the data slightly beyond binomial.

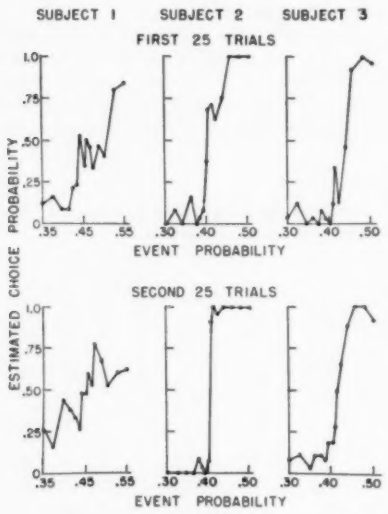

FIG. 3. Estimated choice probabilities vs. event probabilities for S_1, S_2, and S_3 on Payoff Matrix 1 for the first 25 and the second 25 trials. (Each point is based upon 25 observations.)

TABLE 4

P-VALUE COMPARISONS BETWEEN FIRST AND SECOND HALVES OF THE DATA

S	Matrix					
	1	2	3	4	5	6
1	>.05	>.05	.01–.02	>.05	.02	<.01
2	>.05	.01	<.01	>.05	.01	>.05
3	<.01	<.01	.02	.02–.05	.05	.02

Note.—P values are based on Wilcoxon matched-pairs signed rank test.

On the basis of the preceding arguments, we are inclined to view the data as supporting the experimental hypothesis that the choice probability for certain pairs of gambles is a step function of the event probability.

It will be recalled that in contrast to matrices, such as 1–3, in which one row dominates the other, no corresponding theoretical prediction about the response probabilities is known for matrices, such as 4–6, in which the two largest payoffs lie on a diagonal. This may just be our ignorance or it may be that the theory is simply noncommittal about these matrices. In any event, the data fail to suggest any behavioral differences. If the predicted steps actually exist, they also appear when neither row dominates the other; if not, then neither class of matrices exhibits steps.

Furthermore, there is no suggestion that the sign of the payoffs differentially affects the qualitative features of the behavior.

Finally, we turn to the other class of data recorded, response times. It is often held that response times, although

ill understood, are closely related to response probabilities, and so it is conceivable that they too might exhibit discontinuities. The means and variances of the response times were plotted as functions both of the response probabilities and of the event probabilities for each *S*-matrix pair. These plots are not presented because, so far as we could see, there is no consistent picture. Much of the possible detail is masked by variability, and those trends that do show through differ both within *S*s and within matrices. Probably one cannot expect to make subtle uses of response times in experiments, such as this one, where no strong external time pressures are imposed.

Summary

In this experiment, each *S* chose repeatedly between pairs of gambles. His payoff on each trial was determined by his choice together with the outcome of a chance event which was completely independent of his choice. Our purpose was to test the mathematically derived prediction that, when certain inequalities are satisfied by the payoffs, *S*'s choice probability is a step function of the event probability. As an alternative hypothesis we selected the intuitively plausible one that the function relating these two probabilities is ogival. Although qualitatively these hypotheses contrast sharply, various a priori considerations make it difficult to discriminate between them.

Data were collected from 5 *S*s. Each was given six payoff matrices with 15 events per matrix and 50 observations per event. Two *S*s exhibited almost perfectly discontinuous functions, and so they were (trivially) consistent with the step function hypothesis and only consistent with the ogival hypothesis in its limiting form. The other 3 *S*s produced what can best be called lumpy functions. Arguments were adduced to suggest that neither variability coupled with an underlying ogive nor abrupt shifts in the location of an ogive are sufficient to explain the data. No other ways have been suggested whereby an ogive could generate data having plateaus.

The tentative conclusion is that the data support the step function hypothesis more strongly than the ogival one, and that indirect confirmation is given to the assumptions underlying the prediction.

References

Anderson, C. D. The constant-ratio rule as a predictor of confusions among visual stimuli of brief exposure duration. Technical Note, 1959, Hearing and Communication Laboratory, Indiana University.

Chipman, J. S. Stochastic choice and subjective probability. In Dorothy Willner (Ed.), *Decisions, values, and groups*. New York: Pergamon, 1960. Pp. 70–95.

Clarke, F. R. Constant-ratio rule for confusion matrices in speech communication. *J. Acoust. Soc. Amer.*, 1957, **29**, 715–720.

Coombs, C. H. On the use of inconsistency of preferences in psychological measurement. *J. exp. Psychol.*, 1958, **55**, 1–7.

Coombs, C. H. Inconsistency of preference as a measure of psychological distance. In C. W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories*. New York: Wiley, 1959. Pp. 221–232.

Davidson, D., & Marschak, J. Experimental tests of a stochastic decision theory. In C. W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories*. New York: Wiley, 1959. Pp. 233–269.

Davidson, D., Suppes, P., & Siegel, S. *Decision making*. Stanford: Stanford Univer. Press, 1957.

Luce, R. D. A probabilistic theory of utility. *Econometrica*, 1958, **26**, 193–224.

Luce, R. D. *Individual choice behavior: A theoretical analysis*. New York: Wiley, 1959.

Luce, R. D., & Raiffa, H. *Games and decisions*. New York: Wiley, 1957.

Mosteller, F., & Nogee, P. An experimental measurement of utility. *J. pol. Econ.*, 1951, **59**, 371–404.

Papandreou, A. G. A test of a stochastic theory of choice. *U. Calif. Publ. Econ.*, 1957, **16**, 1–18.

Ramsey, F. P. *The foundations of mathematics*. New York: Harcourt, Brace, 1931.

von Neumann, J., & Morgenstern, O. *Theory of games and economic behavior*. (2nd ed.) Princeton: Princeton Univer. Press, 1947.

(Received October 17, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 50–56

EXTINCTION EFFECTS FOLLOWING NONDIFFERENTIAL REINFORCEMENT OF AN IRRELEVANT STIMULUS[1]

SALLY E. SPERLING[2]

University of Michigan

In any discrimination learning situation, there are other stimuli present on each experimental trial besides the positive and negative discriminanda that *E* manipulates. These are usually called "extraneous" or "irrelevant" stimuli. Several investigators, e.g., Lawrence (1949, 1950), Babb (1956), and Jeeves and North (1956), have placed two experimentally manipulable irrelevant stimuli into discrimination learning problems with rats and investigated learning with respect to these stimuli. The latter two studies, both of which utilized simultaneous stimulus presentation and choice responses, agreed in showing that *S*s who received initial training during which each of the irrelevant stimuli appeared equally often with the positive and negative discriminanda did not learn a test problem, where one of these stimuli is now positive and the other negative, any faster than control *S*s who received initial training without the irrelevant stimuli being present.

In test learning, where the stimuli were presented together on each trial, it is not possible to determine why the experimental groups did not differ from the control groups. This might be either because their previous reinforcement history with respect to the stimuli was ineffective in increasing the probability of response in the presence of these stimuli or because the increased probability of response was equal for the two stimuli since both of them were treated alike in terms of pairings with the positive and negative relevant stimuli and in terms of reinforcement.

That the latter alternative may hold is suggested by the consideration that the irrelevant stimuli may be thought of as being under a 50% partial reinforcement schedule during initial training; i.e., they were each associated with reinforcement and with nonreinforcement half of the time that they were present. If this is sufficient to determine partial reinforcement effects, then the consequences of nondifferential reinforcement of an irrelevant stimulus should be like those observed following 50% partial reinforcement (Lewis, 1960).

The purpose of this study was to examine resistance to extinction in the presence of a previously nondifferentially reinforced irrelevant stimulus, where initial training and extinction were given in a runway under single stimulus presentation to eliminate the difficulties posed by choice responses. If the past reinforcement history with respect to the irrelevant stimulus determines the level of extinction performance, then this group ought not to differ from a group which has had 50% partial rein-

[1] This report is based on a dissertation submitted to the Horace H. Rackhan School of Graduate Studies of the University of Michigan in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The experimental work was supported by a grant from the National Science Foundation, (G-9779), to J. D. Birch, and the writer wishes to thank him for his innumerable helpful suggestions.

[2] Now at the University of California, Riverside.

forcement. Both of these groups ought to show more resistance to extinction than groups which have had either 100% reinforcement or no previous exposure of the stimulus present during extinction.

METHOD

Subjects.—The *S*s were 55 male hooded rats, aged 90–110 days at the beginning of experimental training, obtained from the colony maintained by the Department of Psychology at the University of Michigan. Four animals were eliminated during the experiment because of illness.

Apparatus.—The apparatus, described in detail by Birch, Ison, and Sperling (1960), consisted of two elevated platforms separated by a partition. The starting platform was 12 in. long and 3 in. wide. The stimulus platform, 18 in. long and 4 in. wide, was divided into two 9-in. sections by a beaverboard partition. A 6½ in. × 4¼ in. swinging door could be hung from a metal frame directly behind a like-sized aperture cut in the partition; a 6-v. flashlight bulb was mounted on the door, 3¼ in. above the stimulus platform facing the starting platform. An unpainted aluminum food cup could be attached to the terminal end of the stimulus platform.

An infrared light beam obtained from two Worner cells was directed 1 in. over the starting platform, ½ in. from its terminal end. When the light beam was broken by *S*, a 100-w. bulb illuminated the stimulus platform, the flashlight bulb mounted on the swinging door went on, and a GE Standard Electric timer calibrated in .01 sec. was started. When *S* stepped off the starting platform, a normally-off microswitch placed under it was released to stop the clock. With the exception of the floor of the stimulus platform, which was mid-gray floor tile, the entire apparatus was painted flat black.

Stimuli.—Two stimulus platforms longitudinally striped with black and white Mystic tape were used during acquisition training. One of these had stripes 1 in. wide (W); the other had stripes ½ in. wide (N). These two platforms were striped to the swinging door and were covered with longitudinally laid gray tape from the door to their terminal ends. A third platform was completely covered with gray tape. The swinging door was covered with black tape which concealed the mounting of the flashlight bulb. Stimulus platforms could be changed by an arrangement of snap fasteners. During extinction trials the swinging door and its frame were removed. The two stimulus platforms used during extinction were identical to the acquisition striped platforms except that the stripes extended the length of the platforms.

Preliminary training.—There were 14 days of pretraining. The first 6 days were devoted to handling and taming. During the last 8 days *S*s were under 23-hr. deprivation, eating 10 gm. of dog pellets during a 1-hr. feeding period. On the first 5 of these days *S*s were placed on elevated platforms of unpainted pine and allowed to eat from attached food cups. On the last 3 days *S*s received a total of five rewarded trials in the experimental apparatus with an unpainted pine stimulus platform in place. On the two trials of the last day of pretraining, a black painted aluminum door was in place across the stimulus platform.

Experimental groups.—The *S*s were randomly assigned to one of six treatment groups. Five of these received 8 acquisition trials per day for 10 days; the sixth received 8 trials per day for 5 days. For groups wnLD[1] ($N = 9$) and LD ($N = 8$), "light bulb on" (L) was the positive stimulus and "light bulb off" (D) the negative stimulus. For the first of these groups, W and N were each present on half of the positive and half of the negative trials; for the second group, the gray platform was always present. For Groups WNld,a ($N = 10$) and WN ($N = 8$) W was positive and N negative; L and D were each present on half of the positive and half of the negative trials for the former, and D obtained on all trials for the latter. Group wn ($N = 8$) was presented with W and N under the D condition; each of the two striped stimuli was associated with a 50% partial reinforcement schedule. Group WNld,b ($N = 8$) had the same stimuli as Group WNld,a but received only 40 acquisition trials.

All *S*s received 16 extinction trials on the day after their last acquisition day. The light bulb, swinging door, and frame were removed from the apparatus for these trials. Thus the illumination in the stimulus compartment was like that obtaining under the D condition during acquisition. For 4 Group LD *S*s the N platform was presented during extinction. The W platform was present for all others.

[1] Upper case letters are used to designate stimuli associated with 100% and 0% reinforcement and lower case letters to designate stimuli associated with 50% reinforcement.

TABLE 1

ACQUISITION AND EXTINCTION CONDITIONS FOR EXPERIMENTAL GROUPS

Variable	Groups						
	wnLD $N = 9$	wn $N = 8$	WNld,a $N = 10$	WN $N = 8$	WNld,b $N = 8$	LD $N = 4$	LD $N = 4$
Acquisition stimuli							
Light on	100%[a]	—	50%	—	50%	100%	100%
Light off	0%	—	50%	—	50%	0%	0%
Wide stripe	50%	50%	100%	100%	100%	—	—
Narrow stripe	50%	50%	0%	0%	0%	—	—
Acquisition trials	80	80	80	80	40	80	80
Extinction stimulus	W	W	W	W	W	W	N

[a] Percentage of reinforcement during acquisition trials.

The experimental design is presented in Table 1.

Experimental procedure.—All *S*s received 8 trials per day, of which 4 were rewarded. Each *S* in each of the three complex stimulus groups had two pairings of each of the four possible combinations of stripe and light conditions each day. Gellermann (1933) orders were used for the schedules of presentation by omitting an initial or terminal pair of alternating trials from the 10-trial schedules. A table of random numbers was used to determine the complex stimulus presentations for those groups that had them. Within the limits of differences in group *N*s and stimuli presented, schedules were matched across groups. The light bulb on the swinging door was unscrewed but left in place for those trials or groups where the L condition was not called for.

The *S*s were given experimental trials in groups of 5 to 7, maintaining a constant 15-min. intertrial interval for both acquisition and extinction. During initial training the food cup was baited with two .125-gm. pellets for rewarded trials and *S* was removed after 20 sec., which was sufficient for consumption of the reward. On nonrewarded trials *S* was detained for 20 sec. with no food cup present. The *S*s were under 21-hr. food deprivation during acquisition training and 19¾-hr. deprivation at the beginning of the extinction day.

All statistical analyses were in terms of logarithms of reciprocal latency × 100 (log RS).

RESULTS

Table 2 shows mean log RS on W trials on the last day of acquisition training for Groups wnLD; WNld,a; WNld,b; wn; and WN, and mean log RS over all trials on that day for Group LD, which did not have stripes present during acquisition. An analy-

TABLE 2

MEAN LOG RS FOR W TRIALS ON LAST DAY OF ACQUISITION AND FOR 15 EXTINCTION TRIALS

Measure	Group					
	wnLD	wn	WNld,a	WN	LD	WNld,b
Acquisition	1.74	2.03	2.03	2.04	1.70[a]	1.66
Extinction	1.36	1.83	1.56	1.12	.88	1.08

[a] Mean log RS over all 8 trials of the last day, since Group LD was not presented with W during acquisition.

sis of variance on the data of the first five groups yielded an F of .85 ($df = 4/38$; $P > .20$) indicating that terminal level of responding in the presence of the stimulus to be presented during extinction cannot be considered different among these groups. A second analysis including Group LD yielded an F of .88 ($df = 5/45$; $P > .20$), indicating that Group LD's terminal response level may be considered equivalent to that of the other groups.

Since all groups showed a marked increase in mean log RS from first to second trials, and since the first trial was not properly an extinction trial, the first trial was omitted from all subsequent analyses. Two different analyses of the extinction data were made. The first concerns log RS during 15 extinction trials, the second concerns mean decrease in log RS in the presence of W from the last acquisition day to the extinction day.

Group wn ran fastest during extinction, followed in order by Groups WNld,a; wnLD; WN; WNld,b; and LD. All groups showed a decrease in log RS over the extinction trials. Table 3 summarizes an analysis of variance on these trials. For the main effects it can be seen that there were group differences in log RS during

TABLE 3

ANALYSIS OF VARIANCE OF LOG RS FOR 15 EXTINCTION TRIALS

Source	*df*	*MS*	*F*
Between *Ss*	50		
Groups (G)	5	15.07	3.71*
Error (b)	45	4.06	
Within *Ss*	714		
Trials (T)	14	.89	2.62**
T × G	70	.24	.70
Error (w)	630	.34	

* $P < .01$.
** $P < .005$.

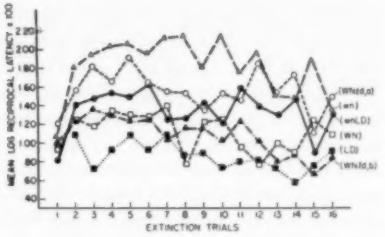

FIG. 1. Mean log RS on 16 extinction trials in the presence of W.

extinction and that there was a significant decrease in response speed over groups. The nonsignificant Groups × Trial interaction suggests that the amount of decrease did not differ among groups.

The rejection of the hypothesis of equal log RS among groups allowed individual comparisons to be made. The means for the groups are presented in Table 2. The difference between the means of Groups wnLD and LD yielded a t significant at the .05 level. All other differences were tested by 95% confidence intervals based on Tukey's method (Scheffé, 1959) to provide descriptive information. Group wn ran faster than Groups LD; WNld,b; and WN.

For the second analysis of the extinction data, mean log RS for each *S* in each of the experimental groups for its 4 W trials on the last acquisition day was subtracted from its mean log RS over the 15 extinction trials. Group LD was omitted from this analysis since these *Ss* did not have W present during acquisition. An analysis of variance performed on the decrease scores indicated that there were differential amounts of decrease among the groups ($F = 3.56$; $df = 4/38$; $P < .05$). Tukey 95% confidence intervals were used to test these differences. Groups wn and

wnLD both showed less decrease than Group WN. The differences between Groups WNld, a and WN and Groups WNld,b and WN approached significance. Each of these differences was tested by means of a 98% confidence interval based on t to provide a posteriori information. Group WNld,a showed less decrease than Group WN; the other difference was not significant.

Discussion

The following interpretation of the results of this study rests upon the assumption that the primary determinant of extinction performance in the various groups was the reinforcement history associated with W during initial training. The stimulus situation during extinction may be considered to be most closely analogous to W and D, and the reasons for omitting consideration of D should be mentioned. These reasons are: (*a*) During extinction the light bulb, swinging door, and frame were removed from the apparatus. The only way, then, in which the extinction situation resembled the D acquisition condition was in the level of illumination of the stimulus platform. (*b*) Groups wn and WN received all of their acquisition training under the D condition, i.e., they were both 50% reinforced in the presence of D. There is some question as to whether the constant level of illumination would come to exercise any control over *S*s behavior, thus the omission of a lower-case "d" from the designations of these groups. (*c*) Analyses of the past reinforcement history of the WD pattern for the different groups or of the combined reinforcement histories of W and of D separately do not lead to any increased explanatory power with respect to these data than that resulting from the consideration of W alone.

If the past reinforcement history with respect to W is considered to be the primary determinant of extinction performance, then, with the exception of Group WNld,a, the results of this study form a reasonable and fairly simple pattern.

Resistance to extinction following nondifferential reinforcement of an irrelevant stimulus during discrimination training on other stimuli in Group wnLD was greater than the resistance to extinction exhibited by Group LD with no previous reinforcement history with respect to this stimulus and equivalent to the resistance to extinction observed in Group wn following 50% reinforcement.

That the first result is not due to disruption of performance in Group LD as a result of the changed stimulus conditions from acquisition to extinction is suggested by the fact that Groups wn and WN both had the same stimulus change and their extinction performance was clearly different, i.e., the partial reinforcement effect is generalizable to a two-stimulus training situation under a constant 15-min. intertrial interval. In addition, Groups WN and LD had different amounts of stimulus change from acquisition to extinction and there was apparently little difference in their extinction performance.

Group WNld,a, which received 100% reinforcement in the presence of W during acquisition training, should have exhibited less resistance to extinction than Groups wnLD and wn and should not have differed from Group WN. The data, however, indicated that a reverse relationship may hold, since Group WNld,a did not differ from Groups wnLD and wn, the 50% reinforcement groups, and showed less decrease than Group WN, the comparable 100% reinforcement group.

The initial training conditions for Groups wnLD, wn, and WNld,a were alike in one respect not shared by Group WN. All of the first three groups had a 50% partial reinforcement schedule present during acquisition training although this schedule was associated with W and N for Groups wnLD and wn and with L and D for Group WNld,a. It has been demonstrated a number of times that resistance to extinction is greater following 50% reinforcement than following

100% reinforcement under conditions where separate groups have been used for this comparison. To the writer's knowledge, this study is the first in which resistance to extinction has been examined following the concurrent administration of 50% and 100% reinforcement schedules. Under these conditions, response strength, as measured by resistance to extinction, seems to be the same in the presence of a previously 50% and a previously 100% reinforced stimulus, suggesting that resistance to extinction in both cases may be some sort of resultant of the administration of two concurrent schedules.

There were no differences between Groups WNld,a and WNld,b under either analysis of extinction data, indicating that Group WNld,a was still in the range of trials where resistance to extinction is positively related to the number of rewarded trials.

There is no evidence in these data for the operation of the mediational process proposed by Lawrence (1949, 1950) to account for the results of test learning following complex stimulus training. He assumed that relevant stimuli become more distinctive during initial training while irrelevant stimuli either do not acquire or lose distinctiveness. New responses are thus more easily learned to relevant than to irrelevant stimuli during test learning through the mediation of acquired distinctiveness. It is not clear whether responses in the presence of more distinctive stimuli ought to be more or less resistant to extinction than responses in the presence of less distinctive stimuli, but a difference between them is clearly implied. In the present experiment, however, no such difference was observable during extinction in the presence of W between Group WNld,a for which W had been a relevant stimulus, and Group wnLD, for which W had been an irrelevant stimulus.

The absence of a strong discriminative response on the initial problem in both of these groups may provide a complicating factor since it could be argued that the modification of initial orders of distinctiveness does not occur until *S* is performing at a high level of correctness on the initial problem. Alternatively, however, it is possible that an analysis of the training conditions used by Lawrence in terms of reinforcement schedules associated with complex patterns of stimuli and position responses may provide an explanation of his data that makes the appeal to mediational processes unnecessary.

Summary

Five groups of rats were given 80 acquisition trials and one group 40 acquisition trials in an elevated runway under single stimulus presentation, using Wide (W) and Narrow (N) striped stimulus platforms and a light bulb, mounted on a swinging door, that was either on (L) or off (D). All *S*s were given 16 extinction trials in the presence of W with the light bulb and swinging door removed from the apparatus.

The results were: (*a*) Group wnLD, for which W had been a nondifferentially reinforced irrelevant stimulus during discrimination training on L and D showed more resistance to extinction than did Group LD which also received discrimination training on L and D but in the presence of a gray platform. (*b*) Group wn for which W had been a 50% partially reinforced stimulus also showed greater resistance to extinction than Group LD and did not differ from Group wnLD. Group wn was more resistant to extinction than Group WN for which W had been the positive discriminandum, demonstrating that the partial reinforcement effect is generalizable to a two-stimulus training situation.

(*c*) Group WNld,a for which W had also been the positive discriminandum, with L and D present as irrelevant stimuli, did not differ in extinction from Group WNld,b, the 40-trial acquisition group, or from Groups wnLD and wn and was apparently more resistant to extinction than Group WN, the comparable 100% reinforcement group.

Except for these last results, the data are consistent with the interpretation that the primary determinant of extinction performance is the past reinforcement history associated with the stimulus presented during extinction, regardless of the previously relevant or irrelevant status of this stimulus.

Thus nondifferential reinforcement of an irrelevant stimulus may be considered to be equivalent to 50% partial reinforcement.

REFERENCES

BABB, H. Proportional reinforcement of irrelevant stimuli and transfer value. *J. comp. physiol. Psychol.*, 1956, **49**, 586–589.

BIRCH, D., ISON, J. R., & SPERLING, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

GELLERMANN, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *J. genet. Psychol.*, 1933, **42**, 207–208.

JEEVES, M. A., & NORTH, A. J. Irrelevant or partially correlated stimuli in discrimination learning. *J. exp. Psychol.*, 1956, **52**, 90–94.

LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–785.

LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–189.

Lewis, D. J. Partial reinforcement: A selective review of the literature since 1950. *Psychol. Bull.*, 1960, **57**, 1–28.

SCHEFFÉ, H. *The analysis of variance.* New York: Wiley, 1959.

(Received October 19, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 57–61

THE ROLE OF DRIVE (TIME STRESS) IN COMPLEX LEARNING: AN EMPHASIS ON PRELEARNING PHENOMENA [1]

R. ERNEST CLARK [2]

State University of Iowa

The efficacy of Hullian drive theory (Hull, 1943; Spence, 1956) for predicting drive effects in complex learning, though well demonstrated by Spence and his colleagues (e.g., Spence, Farber, & McFann, 1956; Spence, Taylor, & Ketchel, 1956), has been controversial because of some antagonistic evidence (e.g., Hughes, Sprague, & Bendig, 1954), and because of the many assumptions typically surrounding the predictions (Hill, 1957). These assumptions have been of three types: (*a*) theoretical assumptions concerning the interactions between such relevant Hullian variables as habit strength, drive, threshold, and oscillatory inhibition; (*b*) coordinating assumptions required in the application of Hullian drive theory to complex learning, e.g., that "manifest anxiety" (Taylor, 1953) is a valid measure of generalized drive; and (*c*) special assumptions to supplement inadequate empirical data regarding the states of specific task variables, e.g., assumed numbers and relative strengths of expected competing responses.

The present study was designed to eliminate many of the special and coordinating assumptions typically used in predictions of complex learning performance by establishing before learning the actual empirical effects of the drive variable on the numbers and relative strengths of competing responses. Given these empirical data, specifically relevant to the present task and sample of *S*s, it was possible to predict early learning performance without presupposing the correctness of Hullian theory, or its applicability to complex learning.

METHOD

Apparatus [3]

The stimulus consisted of a short vertical bar of light resting on a long horizontal line of light. The bar was viewed by *S* as a movable divider that could be made to cut the line into any two segments. Actually, during the experiment proper, only one position of the bar on the line was used. This position was exactly halfway between center and the left end of the line.

The response unit consisted of a stylus and semicircular response slot, with a "ready" hole for the stylus as the origin of the semicircle. Lining the base of the response slot were ¼-in. wide brass contacts which, when touched with the tip of the stylus, allowed for a determination of the spatial response. A head-rest clamping device permitted *S* to see the stimulus, but not the response slot.

Subjects

Thirty-two volunteers from the elementary psychology classes of the State University of Iowa were assigned to a high-drive group and 32 to a low-drive group. Random assignment of *S*s was restricted by the following selection criteria: (*a*) groups were balanced with regard to three levels of *MA* scale scores (Taylor, 1953) to reduce the possibility of disproportional interaction of "manifest

[1] This investigation represents part of a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the State University of Iowa (1958).

[2] Now at the Quartermaster Research and Engineering Command, United States Army, Natick, Massachusetts.

[3] The detailed apparatus has been described elsewhere (Gerjuoy & Bechtoldt, 1955).

anxiety" with the drive conditions; (*b*) 16 *Ss* in each group were men and 16 were women; (*c*) each *S* had a minimum "visual acuity" of 25/20 as measured with the Snellen Eye Chart.

Preliminary Training

The response required of *S* was to duplicate the divisions seen in the stimulus by pushing the stylus into the response slot to a point which would divide the slot appropriately. Six uncorrected practice trials were given, each trial consisting of a 4-sec. silent period, a 2-sec. "ready" tone, and a 4-sec. stimulus duration. It was emphasized that all of the time available should be used to make an accurate response, but that the response to each stimulus should be completed before the stimulus light went off. The 4-sec. stimulus duration used in this phase of the study allowed *Ss* to make rather leisurely responses.

Experiment Proper

The study proper was divided into three sets of 21 trials each, with the reinforcement of selected responses occurring only during the third set of trials. Drive was defined in terms of "time stress," or the time allowed *S* to complete a response. This drive variable has been used successfully before (e.g., Castaneda & Palermo, 1955), and was well suited for the present study since it was easily controlled by varying the duration of the stimulus light.

Trials 1–21 (initial response hierarchies).—During this period all *Ss* responded 21 times to the 4-sec. stimulus, i.e., under minimal time-stress level. The relative frequencies of occurrence of spatial responses made to the stimulus indicated the structure of the competing response hierarchies for the two groups under a constant, and presumably low, drive level.

Trials 22–42 (response hierarchies under different drive levels).—The *Ss* continued responding without interruption for another 21 trials while different levels of time stress were induced by the following conditions: Group 1 (reduced time, high time stress): The stimulus duration was shifted from 4 sec. to a duration giving *S* barely enough time to respond. The actual duration used for each *S* was adjusted on the basis of his mean response speed during the first experimental period. Group 2 (control, low time stress): The *Ss* of this group continued responding to the 4-sec. stimulus. The same low level of time stress was expected to continue.

The response hierarchies obtained during Trials 22–42 showed the specific influences of varied time stress on competing response frequencies just prior to learning.

Trials 43–63 (learning).—During the final 21-trial period, *Ss* continued to respond under varied time stress. However, just before the first learning trial, *Ss* were informed that all correct responses would now be followed by the sound of a chime, and that responses not followed by the chime were to be considered inaccurate.

During learning, half of the 32 *Ss* in each time-stress group learned an easy response and half a difficult response, the relative difficulty of responses being estimated from the previously obtained relative response frequencies. Thus, the easy correct response was selected as *S*'s most probable (dominant) response and the difficult correct response as the very improbable response lying one interquartile range to the right of the dominant response. The interquartile range for each *S* was calculated from the frequency data obtained during Trials 1–21.

Results and Discussion

Effects of Time Stress on Response Hierarchies

The mean frequencies of various spatial responses elicited by the stimulus before (Trials 1–21) and after (Trials 22–42) time stress was varied are shown in Fig. 1. The center "tick marks" in the abscissas of the figure correspond to each *S*'s dominant, or most frequent, response. The marks to the left and right of center refer to the contact regions in the slot immediately adjacent to the dominant responses.

From these data it was possible to predict the early learning performance of high and low time-stress groups for any selected responses. For example, in the present case of the easy (dominant) correct response, the high time-stress *Ss* should have produced more correct responses early in learning than the low time-stress *Ss*. This prediction based upon the data in Fig. 1 was reinforced by a

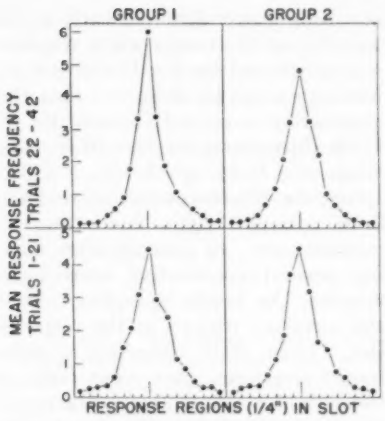

FIG. 1. Hierarchies of competing responses before (Trials 1–21) and after (Trials 22–42) time stress was varied.

statistical analysis of the dominant response frequencies. Variance analyses and subsequent *t* tests indicated that during Trials 1–21 the two groups were equivalent in mean frequencies of dominant responses. When time stress was varied during Trials 22–42, however, a significant increase occurred in the frequency of this response for Group 1 (increased time stress), but not for Group 2 (constant time stress). The means and *SD*s associated with the analyses are shown in Table 1.

Because of the infrequency of the difficult correct response, no comparable statistical analysis was possible. However, since the known prelearning frequency of this response approached zero for both experimental groups, neither group was expected to perform very well early in learning.

Before going on to the learning results, it is of interest to examine the effects of varied time stress upon the overall hierarchies of competing responses, including those responses not actually scheduled for reinforcement

TABLE 1

MEANS AND *SD*s OF FREQUENCIES OF DOMINANT RESPONSES BEFORE (TRIALS 1–21) AND AFTER (TRIALS 22–42) TIME STRESS WAS VARIED

Group	Trials 1–21		Trials 22–42	
	Mean	*SD*	Mean	*SD*
1 increased time stress	4.50	2.18	6.06	2.88
2 constant time stress	4.41	2.05	4.47	1.81

Note.—Critical difference based on $P < .05$ between column means is 1.12, and between row means is 1.03.

during learning. If time stress had functioned multiplicatively, changes in response strengths accompanying increased time stress (Trials 22–42) would have been proportional to the original response strengths (Trials 1–21). Generally, strong responses would have increased in frequency and weak responses decreased.

Figure 2 shows the relationship between original response frequencies and those obtained under the varied time-stress conditions. In such a plot of the response variable, a "no-change" reference may be represented by a straight line with slope of one and origin of zero. Had no changes occurred in the respective frequencies of responses from Trials 1–21 to Trials

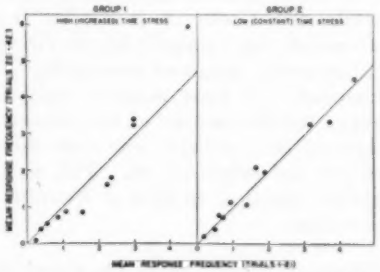

FIG. 2. Relative response frequency during Trials 22–42 as a function of response frequency during Trials 1–21 and time-stress level.

22–42, all of the plotted points would have fallen upon the reference lines.

Considering first the data of Group 1, it is obvious that increased time stress was accompanied by systematic changes in relative response frequencies. These changes appear to be largely consistent with the notion of time stress as a multiplicative drive variable, and consequently, with Hullian drive theory. Stronger responses increased in frequency in proportion to their original relative strengths, and weaker responses decreased in proportion to their original relative weakness. Although the very weak competing responses, initially zero to one in mean frequency, seemed to be unaffected by varied time stress, little confidence may be placed in these data because of the very few scores involved. Even considered collectively, these extremely weak responses represented less than 4/21 of the total response frequency.

It should be made clear at this point that, although the effects of time stress were not contrary to Hullian theory, the present data are not in such form as to support the Hullian multiplicative rule in unequivocal preference to other monotonic functions that may suitably relate drive and habit variables.

It was apparent from Fig. 2 that, while the Group 1 hierarchy was altered during Trials 22–42, the Group 2 hierarchy remained essentially unchanged. A least squares analysis suggested the best-fit of the Group 2 data to be a straight line with slope of .96 and origin of .08. This fit is almost identical to that of the reference line.

Effects of Time Stress on Learning Performance

Learning an easy response.—Figure 3 shows that the predicted superiority of Group 1 over Group 2 early in the learning of the easy correct response was supported by the learning data. Variance analyses indicated that this superiority remained statistically reliable throughout the first 10 learning trials ($F=14.14$, $df=1/150$, $P<.05$). Later, the differences disappeared and both groups began fluctuating in performance. In general, after making several consecutive correct responses, *S*s began sampling across the incorrect regions in the response slot. Then, after numerous nonreinforced responses, they went back to the correct response area. The result of this kind of behavior is obvious in the "dips" and "rises" of the functions in Fig. 3.

Learning a difficult response.—As illustrated in Fig. 3, neither group produced many correct responses early in the learning of the difficult response. An analysis of variance indicated no group differences in correct response frequencies throughout the 21 learning trials. There was, however, a significant trials effect ($F=3.11$,

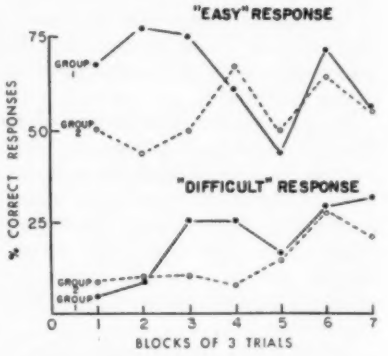

FIG. 3. The percentage of correct responses produced in blocks of three trials by each group in learning the easy and difficult responses. (Group 1 refers to the high time-stress group and Group 2 to the low time-stress group.)

$df = 6/180, P < .01$), suggesting that some learning did take place.

The relative brevity of the learning periods resulted from the present experimental interest in the predictability of *early* learning performance from the *known* prelearning effects of drive. To this end, the present methodology proved to be quite adequate.

Summary

The role of "time stress" (drive) in complex learning and performance was investigated by relating empirical changes in prelearning competing response hierarchies to differences in the early learning performance of high and low time-stress groups. The relative strengths of competing responses were provided before learning by relative frequencies of occurrence of spatial responses to a single stimulus in a motor paired-associate task. The response consisted of dividing a slot with a stylus so that the divisions corresponded with those presented in a stimulus line. When given a number of trials, *S*'s responses to the single stimulus were variable enough to yield a hierarchy of competing responses with differing relative frequencies of occurrence.

The 63 trials of the experiment proper were divided into three periods of 21 trials each, reinforcement of selected responses occurring only in the last period. First, an initial response hierarchy was obtained while all *Ss* responded under the same low level of time stress. Second, changes in the competing response hierarchies were obtained while time stress was varied. Finally, half of the *Ss* in each of two time-stress groups learned an easy response and half a difficult response.

It was found that the effects of time stress (drive) upon prelearning competing response hierarchies were largely compatible with Hullian drive theory. Predictions of early learning performance, based on these known prelearning effects, were supported by the learning data.

References

Castaneda, A., & Palermo, D. S. Psychomotor performance as a function of amount of training and stress. *J. exp. Psychol.*, 1955, **50**, 175–179.

Gerjuoy, H., & Bechtoldt, H. P. *Generalization effects in successive discrimination learning as measured by response similarity.* Final Report, March 8, 1955, State University of Iowa, ONR Contract Nonr 1261-(00).

Hill, W. F. Comments on Taylor's "Drive theory and manifest anxiety." *Psychol. Bull.*, 1957, **54**, 490–493.

Hughes, J. B., Sprague, J. L., & Bendig, A. W. Anxiety level, response alternation, and performance in serial learning. *J. Psychol.*, 1954, **38**, 421–426.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Spence, K. W., Farber, I. E., & McFann, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.

Spence, K. W., Taylor, J. A., & Ketchel, R. Anxiety (drive) level and degree of competition in paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 306–310.

Taylor, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received October 20, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 62–67

THE EFFECTIVENESS OF CONCEPTS AT VARIOUS LEVELS OF AWARENESS

KEITH G. DAVIS AND HARRIE F. HESS[1]

University of Colorado *University of Oregon Medical School*

Throughout the history of psychology there has been continuing interest in phenomena which may be classified under the rubric of "behavior without awareness." It is generally conceded that concepts may be formed without awareness. The recent studies on verbal conditioning (Krasner, 1958) provide contemporary evidence bearing on this point. The issue with which this paper deals is whether or not concepts which operate below awareness can guide behavior efficiently.

One study which has frequently been cited as providing evidence which indicates that unconscious concepts can effectively guide behavior is that of Rees and Israel (1935). Leeper (1951) cites the Rees and Israel experiment as one showing ". . . that concepts may be formed, retained, and used 'unconsciously'." In a recent review, Adams (1957) cites the Rees and Israel experiment as one of the few experiments in one major category providing unequivocal, positive evidence of behavior without awareness. Other authors refer to the Rees and Israel study in a similar manner (Hebb, 1958; Leeper & Madison, 1960; Woodworth & Schlosberg, 1954).

In the Rees and Israel experiment, *Ss* who *were not* aware of a system for solving anagrams used the system about as effectively as other *Ss* who *were* aware of the nature of the system. This finding that awareness is essentially unrelated to the adaptive use of concepts contradicts contemporary theoretical formulations (Dollard & Miller, 1950; Kelly, 1955; Rogers, 1951; Rotter, 1954), which stress the importance of awareness. This contradiction, taken together with the facts that Rees and Israel used only 10 *Ss*, and that their observations of awareness were merely anecdotal, suggested that the pertinent part of the Rees-Israel study be replicated.

[1] Now with the Nevada State Department of Health, Las Vegas, Nevada.

Rees and Israel (1935) established a set (or, according to subsequent interpretations, they taught their *Ss* a concept) by training them on a series of 15 anagrams, each of which had only one possible solution. The solutions to the training anagrams could be found by rearranging the letters according to a certain letter order. A second series, consisting of 15 multiple-solution anagrams (i.e., anagrams which could be solved by the letter-order pattern or by alternative letter rearrangements), was then presented. The strength of the induced set was measured by counting the number of anagrams in the second series solved according to the letter-order pattern. All of Rees and Israel's experimental *Ss* were highly influenced by the "set." They solved 95% of the multiple-solution anagrams according to the letter-order pattern as compared with 47% for 10 control *Ss*. The remarkable aspect of the findings was that 6 of the 10 experimental *Ss* reported no conscious awareness of having used a pattern to guide their solutions, although 91% of their solutions were according to the pattern. Rees and Israel (1935) concluded that, "The most significant feature of this experiment was its demonstration that the set could operate at such a high level of effectiveness without the subject's being aware of its existence" (p. 15).

The present paper is a report of a replication of the Rees and Israel experiment which includes: (*a*) a larger and more heterogeneous sample, (*b*) a more systematic method of assessing awareness, and (*c*) an additional measure of the effectiveness of the concept. The research hypothesis, derived from contemporary theorizing in the field of personality, was that contrary to the results of the Rees and Israel study, awareness of the pattern or concept would be highly related to effectiveness of the pattern in guiding solutions to anagrams.

METHOD

Subjects.—Seventy-four experimental and 24 control *Ss* were studied. The 74 experimental *Ss* consisted of four different groups; (*a*) 24 introductory psychology students, (*b*) 18 firemen, (*c*) 10 psychology graduate students, (*d*) 22 high school students from the upper 25% of a senior class. The control group consisted of 24 introductory psychology students.

Procedure.—*Ss* were seen individually. The procedure was essentially identical to that used by Rees and Israel (1935). The anagrams were typed in lower-case letters and presented singly by employing a cardboard strip with a small window. The *S* was required to continue on each item until a solution fitting the mechanical requirements was obtained or until 90 sec. had elapsed. At the end of 90 sec., *S* was required to leave the item and go on to the following one. The *E* also timed each solution.

Experimental design.—The design of the experiment was the same as that used by Rees and Israel (1935, p. 14). All *Ss* were required to solve 10 practice anagrams with haphazard letter orders before beginning the experiment proper. Following the practice series, all *Ss* were required to solve 30 anagrams. For the experimental group, the 30 anagrams consisted of two series of 15 anagrams each: the training series and the test series. The 15 anagrams of the training series were single-solution anagrams with letters of the words rearranged in a 54123 pattern (SCENE is the anagram ENSCE). The test series consisted of 15 multiple-solution anagrams soluble both by the letter-order pattern, and by one or more alternative arrangements.

The 30 anagrams for the control group were made from the same words as those used with the experimental group. The 15 multiple-solution anagrams were presented to the control group in form identical to the form of presentation to the experimental group, but they were interspersed with the single-solution anagrams. The 15 single-solution anagrams had letters arranged in haphazard order. The actual list of anagrams used and the order of presentation can be found in Rees and Israel (1935, pp. 24–25).

The reader will note that the method of presentation of the "test series" (i.e., the 15 multiple-solution anagrams) was different for the experimental and control groups. Consequently, differences between experimental and control groups on the response measures cannot confidently be attributed to the effects of the training series. This defect in the original experimental design is inconsequential to the purpose of the present paper, which is concerned only with differences between *Ss* within the experimental (trained) groups.

Measures of the influence of training.—Two measures were used to determine the effectiveness of the letter-order pattern in guiding the solutions of the experimental groups on the test anagrams. The first measure was the same as used by Rees and Israel (1935), viz., the number of multiple-solution anagrams solved according to the letter-order pattern by each *S*. The second measure, response latency, was the mean time required for each *S* to solve the multiple-solution anagrams.

The measurement of awareness.—Rees and Israel (1935) used a postexperimental interview to obtain a dichotomous measure of the extent to which *Ss* were aware of the letter-order pattern. To overcome the disadvantages of the interview and to provide a graduated, rather than a dichotomous measure of awareness, the present experiment utilized a 4-point awareness rating scale similar to that devised by Matarazzo, Saslow, and Pareis (1960). In this scale, awareness ratings were based upon *Ss*' written replies to three open-ended sentences mimeographed on two sheets of paper, presented separately. The sentences were: (Sheet 1) "The purpose of this experiment is . . .," and "My evidence for this is . . .," and (Sheet 2) "The method I used was . . . ." Judges independently rated the written responses of *Ss* on a 4-point rating scale according to predetermined criteria of awareness patterned after those used by Matarazzo, Saslow, and Pareis (1960).

The reliability of the awareness scale was determined by having four judges independently rate the written replies of the 18 *Ss* in the firemen group. An index of order association (Senders, 1958) was calculated between each pair of raters. The indices were found to be .96, .98, .98, .98, 1.00, and 1.00 ($P < .001$ for an index of .96), indicating a satisfactory reliability of rating. The rater with the highest agreement with other raters completed the awareness ratings for the remainder of *Ss*. All ratings were made without knowledge of the performance of the *Ss* on the anagram series.

RESULTS

Effectiveness of the training procedure.—Table 1 presents the mean

TABLE 1

EFFECTIVENESS OF TRAINING PROCEDURE

Group	*N*	Mean Number of Prescribed Solutions		Mean Latency in Seconds	
		Mean	*SD*	Mean	*SD*
Replication study:					
Controls (introductory psychology students)	24	9.08	2.30	22.76	8.47
Experimental groups:					
Introductory psychology students	24	14.38	1.55	5.25	7.24
Graduate students	10	14.00	1.63	4.34	3.27
Firemen	18	10.61	3.42	21.55	19.18
High school students	22	12.82	2.75	11.02	11.84
Combined groups	74	12.95	2.86	10.81	13.72
Original Rees and Israel study:					
Experimental group	10	14.19[a]		Not Reported	
Controls	10	7.10[a]		Not Reported	

[a] Estimated from data presented by Rees and Israel (1935).

scores of the experimental and control groups on the two response measures. It can be seen that the combined experimental (trained) group showed (*a*) a larger number of anagrams in the test series solved according to the prescribed pattern, and (*b*) shorter latencies than the control group ($P < .001$, Mann-Whitney U Test). Table 1 also presents estimates of the mean number of prescribed solutions for the experimental and control groups of the original Rees and Israel (1935) study. It can be seen that these scores for *Ss* of the original experiment are generally similar to those for the comparable experimental groups (the graduate student groups and the introductory psychology student groups) and the control group of the present study. Thus, it may be concluded that the results of the

TABLE 2

NUMBER AND PERCENTAGE OF *Ss* IN EACH AWARENESS CATEGORY

Awareness Rating	Individual Experimental Groups								Combined Experimental Groups		Control Group (Introductory Psychology Students)		Rees & Israel Experimental Group[a]	
	High School Students		Introductory Psychology Students		Graduate Students		Firemen							
	N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%
Level 1 (unaware)	5	23	1	4	0	0	5	28	11	15.5	17	71	6	60
Level 2	1	4.5	0	0	0	0	4	22	5	6.5	0	0	3	30
Level 3	4	18	6	25	3	30	3	17	16	21.5	7	29	0	0
Level 4 (fully aware)	12	54.5	17	71	7	70	6	33	42	56.5	0	0	1	10
Total	22	100	24	100	10	100	18	100	74	100	24	100	10	100

[a] Estimated from data presented by Rees and Israel (1935).

replication study are similar to those of the original study with respect to the effect of the training procedure on roughly comparable subject samples.

Degree of awareness of the letter-order pattern.—Table 2 presents the distributions of awareness ratings of both experimental and control *Ss* of the present experiment, as well as estimated awareness ratings of Rees and Israel's (1935) original experimental group. The awareness ratings of the Rees and Israel group were made by the present authors from brief descriptions included in the original Rees and Isra[illegible]ort, and must be considered wit[illegible]on.

In Table 2 it can be seen that 84.5% of the combined experimental *Ss* of the present study attained some degree of awareness of the letter-order pattern, whereas only 40% of Rees and Israel's experimental *Ss* gained any awareness. Presumably part of this difference is attributable to the more sensitive method of assessing awareness which was used in the present study.

The relation between awareness and the response measures.—Since the two college student experimental groups (introductory psychology and graduate students) were so uniformly high in awareness (Table 2) it was impossible to examine the relationship in these groups between awareness and response measures. However, the awareness ratings and response measures for firemen and high school students display a wider variation, thus permitting a test of the experimental hypothesis. The statistic used was the index of order association. All indices for these two experimental groups were highly significant ($P < .001$) as shown in Table 3. A definite, positive relationship is evident for the firemen and high school groups between awareness and number of prescribed

TABLE 3

RELATIONSHIP BETWEEN AWARENESS RATINGS AND RESPONSE MEASURES

Group	*N*	Index of Order Association		*z* Score	
		Prescribed Solutions	Latency	Prescribed Solutions	Latency
Firemen	18	.73	−.40	7.96*	4.12*
High school students	22	.77	−.83	8.80*	9.20*
Combined experimental groups	74	.77	−.70	28.7*	24.8*

* $P < .001$.

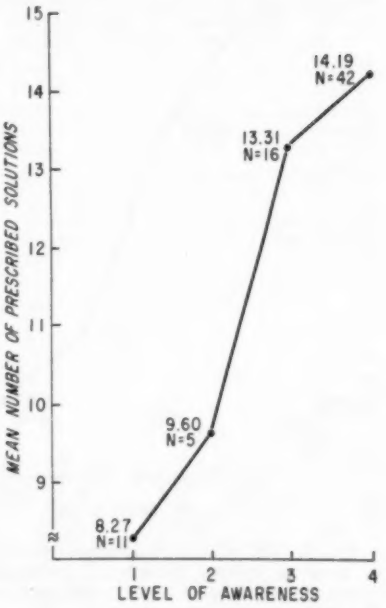

FIG. 1. The relation of awareness ratings to number of test-series anagrams solved according to the letter-order pattern. (Combined experimental groups, $N = 74$.)

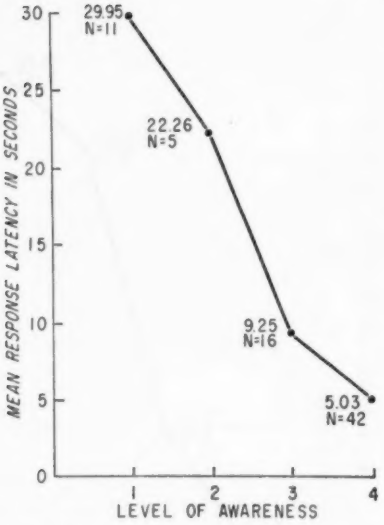

FIG. 2. The relation of awareness ratings to response latency in solving test-series anagrams. (Combined experimental groups, $N = 74$.)

solutions on the test series, as hypothesized. Also, as expected, there is a significant inverse relationship between awareness ratings and response latency.

The last row in Table 3 indicates that the data, when grouped, show the same results as did the separate analysis of the firemen and high school *Ss*. Awareness ratings are highly related to both measures of response efficiency ($P < .001$). Figures 1 and 2 depict these relationships graphically.

DISCUSSION

The results of the present replication differ markedly from those of the original study with respect to the relationship between awareness of the letter-order pattern and the extent of use of that pattern in solving anagrams. Among the Rees and Israel (1935) experimental *Ss*, essentially no relation was found between awareness of the pattern and the influence of that pattern on solutions to the anagrams; in the experimental groups of the present study, the relation between awareness and usage of the pattern was high and significant. In the broad range of *Ss* sampled in the present study, not 1 *S* produced results which suggest an effective, unconscious concept. Thus, the present replication strongly contradicts the Rees and Israel (1935) findings, and in so doing finds itself in agreement with contemporary theoretical formulations concerning the relationship between awareness and adaptiveness of behavior (Dollard & Miller, 1950; Kelly, 1955; Rogers, 1951; Rotter, 1954). The differences found in the relationship of awareness and effective behavior are thought to be due to: (*a*) the improved measure of awareness used in the present study, and (*b*) the larger and more heterogeneous sample used.

SUMMARY

This experiment was a replication of the oft-cited study by Rees and Israel (1935) which reported that 6 out of 10 (60%) of experimental *Ss* developed and used highly effective concepts that were below the level of awareness. In the replication more *Ss* were used than in the original experiment, and a more systematic method of assessing awareness was employed. None of the *Ss* in the replication study developed efficient concepts at a level below awareness. Instead a close relation between awareness and anagram solving efficiency was noted, contrary to one of the main implications of the Rees and Israel study.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 383–405.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

HEBB, D. O. *A textbook of psychology.* Philadelphia: Saunders, 1958.

KELLY, G. A. *The psychology of personal constructs.* New York: Norton, 1955.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

LEEPER, R. Cognitive processes. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 730–757.

LEEPER, R., & MADISON, P. *Toward understanding human personalities*. New York: Appleton-Century-Crofts, 1960.

MATARAZZO, J. D., SASLOW, G., & PAREIS, E. N. Verbal conditioning of two response classes: Some methodological considerations. *J. abnorm. soc. Psychol.*, 1960, **61**, 190–206.

REES, H. J., & ISRAEL, H. E. Establishment and operation of mental sets. *Psychol. Monogr.*, 1935, **46**(6, Whole No. 210).

ROGERS, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.

ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

SENDERS, V. L. *Measurement and statistics*. New York: Oxford Univer. Press, 1958.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Wiley, 1954.

(Received October 21, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 68–73

SIZE-CONSTANCY JUDGMENTS AND PERCEPTUAL COMPROMISE

V. R. CARLSON [1]

National Institute of Mental Health

A generally accepted formula for expressing experimental size-constancy results is the Brunswik (or Thouless) ratio, which reflects the extent to which an obtained size-judgment departs from retinal angular size relative to the difference between a correct retinal value and a correct physical-size value (Brunswik, 1956, p. 21). Although the formula itself allows for the occurrence of numerical ratios outside this range, the derivation and intent of the formula are clearly based on the conception of two essentially limiting values between which perceived size normally falls (Brunswik, 1933, pp. 387–411; Thouless, 1931, pp. 343–344).

Empirically this range seems to cover size-judgments more or less adequately. In the nonreduction situation instructions to judge on the basis of apparent size typically produce average values which fall short of perfect size-constancy ("underconstancy"); analytic, or projective, instructions lead to average values considerably closer to, but still larger than, the lower limit corresponding to a correct retinal value (Brunswik, 1956, pp. 67–69; Gilinsky, 1955; Holaday, 1933; Joynson, 1949; Singer, 1952). These results lend credence to a general view that perceived-size is a compromise between retinal image-size and physical object-size (Brunswik, 1956, pp. 47–48; Osgood, 1953, p. 276; Woodworth & Schlosberg, 1954, p. 483).

[1] The author is greatly indebted to Eugene Tassone for assistance in the collection and analysis of the data.

Instructions to judge in terms of objective, physical-size, however, are likely to result in ratios somewhat greater than 1.00 ("overconstancy"). The average discrepancy in this direction is not usually as great as the deviation from 1.00 in the opposite direction obtained in a projective match, and overconstancy has not seemed to be of equal theoretical importance. It has become increasingly clear, however, that overconstancy is not a rare phenomenon which occurs only under very special circumstances. Its existence is as firm as the existence of underconstancy and cannot be dismissed simply as error (Carlson, 1960; Gilinsky, 1955; Holway & Boring, 1941; Jenkin, 1957; Smith, 1953).

The general hypothesis proposed here is that the principle underlying the Brunswik and Thouless ratios is incorrect. Rather than size-constancy representing a limiting "pole of intention" for size-judgment (Brunswik, 1956, pp. 98–99), it represents the basic functional relationship between object-size and perceived-size. It is the natural neutral point from which size-judgment starts, and substantial deviations from this value can be produced in *either* direction by variations in experimental conditions.

In particular the present experiment is concerned with determining whether overconstancy will occur to as great an extent with appropriate instructions as does underconstancy with analytic instructions. It was previously hypothesized that the perspective attitude operates to produce

overconstancy with objective-size instructions (Carlson, 1960). It follows that if the perspective attitude is made explicit in instructions to *S*, the effect should be enhanced. At the same time there is no basis either in the immediate visual stimulation or in logic for such perspective instructions to cause any effect different from objective-size instructions. If a difference does occur, therefore, it must be due to an attitudinal bias on the part of the observer.

Method

Subjects.—The *S*s were 32 junior college students, 23 men and 9 women, who were paid for serving as *S*s. They were divided into four equal groups, treated exactly the same except for the order in which they received the various instructions. One *S* could not understand the perspective-size instruction (see below) and was not used in the experiment.

Procedures.—Each *S* was tested individually on six different occasions with several days to 1 week intervening between successive sessions. Group 1 received objective-size instructions on Days 1 and 2, perspective-size instructions on Days 3 and 4, and apparent-size instructions on Day 5. Group 2 followed the same schedule but with the order of the objective-size and perspective-size instructions reversed. Groups 3 and 4 were comparable to Groups 1 and 2, respectively, except that the apparent-size instruction occurred on Day 1 instead of Day 5. The apparent-size judgments were included for comparison and correlational reasons secondary to the main purpose of the experiment.

All *S*s received instructions to match for projective-size on Day 6. This task was reserved until last, because it specifically defines the required size-estimate as something different from a size-constancy judgment and would very likely influence the results in the other conditions, none of which, in and of themselves, require any deviation from size-constancy. Finally, following the projective-size matches, *S* was asked to make two distance estimates. In the first, he indicated the subjective location of the midpoint between himself and the apparatus of the standard triangle. In the second, he adjusted the slider on a 40-cm. model to duplicate the distance of the variable triangle relative to the distance of the standard triangle.

Groups 1 and 2 were tested concurrently, as were Groups 3 and 4. The testing of Groups 3 and 4, however, was not started until Groups 1 and 2 had been completed.

In each condition *S* was required to adjust the size of a variable triangle relative to the following altitudes of a standard triangle: 76, 117, and 159 mm., and four values on either side of 117 mm. spaced in steps of 3.2 mm. All values of the standard were presented randomly twice each, requiring a total of 22 trials at each session.

Apparatus.—The triangles were white and isosceles with an altitude-to-base ratio of 1.39. Each was mounted on an apparatus providing continuous variation in size of the triangle against a 30 × 36 in. black felt background. The standard triangle was located at a distance of 40 ft. from *S*, the variable triangle at 10 ft. and at an angle of 20° to the line of sight to the standard. Each apparatus was placed on a table, so that both triangles were approximately at *S*'s eye level. The *S*'s control knob was mechanically coupled to the variable triangle.

Testing was carried out in a well-lighted room similar to a 10 × 50 ft. corridor except for the presence of numerous miscellaneous objects.

Instructions.—The *S* was initially told that the standard triangle would be changed in size from trial to trial at random, that the variable would always be set for him at a randomly determined starting point, and that he could adjust the variable back and forth until he was reasonably satisfied with the match but not to worry unduly over any single setting. Natural, free binocular regard was specified, and *S* was not allowed to utilize devices such as closing one eye or gauging visual extent by sighting between the fingers. The specific instructions were:

> *Apparent-size.*—Adjust the variable so that it looks equal to the standard in apparent visual size. It may also be equal in actual physical size at that point, or it may not—we are not concerned about that. Try to adjust the variable so that it appears equal to you visually, whether you think it is equal in actual size or not.
>
> *Objective-size.*—Adjust the variable so that it is, as best you can judge, equal in actual physical size to the standard, so that if you were to measure them both with a ruler they would measure the same size. They may also look equal to you in apparent visual size, or they may not—we are

not concerned about that. Try to adjust the variable so that you think it is the same actual size as the standard whether it appears equal to you visually or not.[2]

Perspective-size.—As you look down a road or a set of railroad tracks, the edges of the road or the rails appear to come together in the distance. It is that degree of apparent convergence that we want you to try to duplicate here. Set the near triangle so that if the two triangles were lined up, lines extending from the corners of the near triangle, through the corresponding corners of the far triangle, would seem to come together in the distance. Make this setting so that the apparent convergence of these lines would be the same as the apparent convergence of railroad tracks or the edges of a sidewalk or road which extends away from you.

Some repetition and paraphrasing of this instruction was usually required, such as substituting "way in which the lines or railroad tracks seem to come together" for "apparent convergence." The *E* did not provide any sketches or other illustrations.

Projective-size.—The *S* was instructed to adjust the variable triangle so that its subtended angle would be equal to the angle subtended by the standard triangle. The *E* demonstrated that the image of a small extent close to the eye can encompass the visual image of a more distant, large object and illustrated the geometry of the problem by pencil sketches. As much explanation as necessary was given to ensure that *S* clearly understood the result he was to try to achieve in terms of equality of angular, projective-size. Any statement, however, to the effect that object-size should be proportional to distance was carefully avoided.

Results

The ratio of the physical size of *S*'s setting of the near, variable triangle to the physical size of the far, standard triangle was computed for each trial. The average of these ratios for the three main sizes of the standard (76, 117, and 159 mm.) constituted *S*'s score for a given day. A ratio greater than 1.00 indicates "overconstancy," while a ratio less than 1.00 indicates "underconstancy." This ratio is simply that between the sizes of the variable and standard triangles, not a Brunswik or Thouless ratio.

Scores for perspective-size and for objective-size were each obtained on 2 different days in order to evaluate the reliability of the perspective-size judgments. This reliability ($r_{11} = .72$) was reasonably comparable to that ($r_{11} = .85$) for the objective-size judgments (both values based on the entire group of 31 *S*s). The means for the first and second scores were similar in both cases, and the two scores within each condition have been averaged together for all further analyses.

There was no demonstrable effect of order of receiving instructions with respect to whether *S*s received the perspective size instruction prior to the objective-size instruction or vice versa. The *S*s (Groups 3 and 4) who received the apparent-size instruction first, however, clearly manifested a lower mean value for that instruction than those (Groups 1 and 2) who received it later in the schedule (Table 1). This was an unexpected result and requires further comment (see below). But no other significant within-conditions difference occurred among these groups, so the several groups have been combined into a single group of 31 *S*s for all comparisons not involving the apparent-size scores. The statistical significance test employed in each case was a *t* test between correlated means.

The mean objective-size ratio was greater than 1.00, and the mean perspective-size ratio was, in turn, greater than the objective-size value. The projective-size mean was less than 1.00. These differences were all highly significant ($P < .001$). The

[2] The rationale for these apparent-size and objective-size instructions has been given previously (Carlson, 1960).

deviations of the perspective and projective ratios from 1.00 were compared, and the mean difference between the two distributions was clearly insignificant. On the average, *S*s in the perspective situation produced a size ratio as much greater than 1.00 as they did less than 1.00 in the projective situation. The r between the perspective and objective judgments was .47 ($P < .01$); that between the perspective and projective judgments, −.36 ($P < .05$). These questions were the primary ones of the experiment, and the results are unequivocal with respect to them.

A number of previous groups of *S*s have been tested in the present stimulus situation with the same apparent-size instruction. The mean ratios for these groups have been near 1.00, and none has been significantly less than 1.00. Thus, it is the apparent-size mean for Groups 3 and 4 (Table 1) which is anomalous and not that of Groups 1 and 2.

In trying to determine the reason for this discrepancy, it was ascertained that many of the *S*s in Groups 3 and 4 were personally acquainted with individuals who had been *S*s in Groups 1 and 2. It is possible that they had been alerted to the analytic, projective approach to begin with, and it is unfortunate that all the *S*s were not tested with the other instructions prior to testing any with the projective instruction. If this possibility is correct, there should be a correlation between the apparent-size and projective-size judgments for Groups 3 and 4. This correlation was, in fact, .69 ($P < .01$). Whatever the reason for it, these *S*s tended to respond to the apparent-size and projective-size instructions similarly. The corresponding correlation coefficient for Groups 1 and 2 was .37 ($P > .05$).

The remaining intercorrelations among the various conditions, including the distance judgments, were also computed. None reached a statistically significant level, except that the two kinds of distance judgment (midpoint between *S* and standard, and ratio of variable to standard distances) were significantly related to each other ($r = -.64$, $P < .01$).

The nine values of the standard triangle spaced closely around the middle value were utilized to obtain a measure of within-*S*s precision, or sensitivity to relatively small differences in the standard. The rank-order correlation coefficient between these nine values and *S*'s corresponding responses to them was calculated for each *S* for each condition. These coefficients averaged across *S*s were .74, .80, .76, and .76 for the perspec-

TABLE 1

AVERAGE RATIO OF *S*'S SETTING OF NEAR, VARIABLE TRIANGLE TO PRESENTED SIZE OF FAR, STANDARD TRIANGLE

Groups[a]	Measure	Size Instruction			
		Perspective	Objective	Apparent	Projective
1 and 2 ($N = 16$)	Mean	1.43	1.19	1.02	0.62
	SD	0.29	0.17	0.18	0.20
3 and 4 ($N = 15$)	Mean	1.41	1.22	0.79	0.52
	SD	0.23	0.19	0.18	0.17
	P[b]	*ns*	*ns*	<0.01	>0.10

[a] Groups 1 and 2 received the apparent-size instruction after having completed the perspective and objective judgments; Groups 3 and 4 received the apparent-size instruction first.
[b] Probability associated with the t value for the difference between uncorrelated means.

tive, objective, apparent, and projective conditions, respectively. They are all significantly different from zero but do not vary significantly among themselves. Hence there is no reason to suspect that *S*s performed the task any more or less carefully under one instruction than another.

Discussion

The assumption made in the present theoretical argument is that the perspective attitude (Gibson, 1952) is a crucial determinant in specifying a psychologically adequate definition of size in the size-constancy situation. The normal, adult *S*, having lived in a culture which widely uses perspective as a pictorial and communicative representation of object distance, comes to the laboratory already equipped with this attitude. Given this response bias, overconstancy results if *S* is oriented toward judgment in terms of objective-size; underconstancy, if he is oriented toward apparent-size, provided the latter is understood to be something different from objective-size (Carlson, 1960, 1962). The apparent-size instruction used here was specifically designed to eliminate this attitudinal bias, resulting in perfect constancy, at least on the average.

By hypothesis, the perspective instruction devised for the present experiment makes more explicit the bias presumed to be already operating in the objective-size situation. Logically this instruction defines a size match which is no different from an objectively equal match. Since there is a clear increase in the overestimation of size, however, some subjective attitude or assumption must be involved. If it is the same subjective factor present in the objective-size condition, the two sets of judgments should be correlated. The correlation obtained ($r = .47$) was clearly significant.

Similarly, in this view, instructions to match in terms of projective-size make the bias more explicit in the opposite direction. At the same time, this instruction defines a geometrical result which is clearly different from size-constancy and which could be accomplished simply by producing a size ratio which is the same as the distance ratio involved. Some *S*s do seem to try to go about the task this rationally, while others do not. Thus, there is an additional factor which precludes a high correlation with the perspective-size judgments. The lower magnitude of the correlation ($r = -.36$) is therefore reasonable.

Whether this account of the operation of instructions is correct or not, however, the notion that perceived-size is a compromise between physical object-size and retinal image-size does not encompass the occurrence of overconstancy. The further demonstration that overconstancy can occur to as great an extent as underconstancy supports the proposition that size-constancy represents the natural midpoint, rather than one extreme, of the range of experimentally-obtained relationships between perceived-size and object-size. If this hypothesis should prove to have sufficient generality, it suggests two further important implications. The first is that the primary requirement for an adequate theoretical account of size perception is a formulation which specifies invariant perceived-size with varying distance, not increasing or decreasing perceived-size with increasing distance. The second implication is that purported effects of motivational or other personological factors on perceptual constancy would be more adequately formulated as judgmental response biases than as effects upon the perceptual process itself.

Summary

A "perspective-size" instruction designed to make explicit the response bias postulated to be present in size-constancy judgments of objective-size clearly produced a greater degree of overestimation than objective-size instructions. The difference from a physically correct size-match was as great as that obtained in the opposite direction when *S*s judged in terms of projective-size. Perspective-size was positively correlated with objective-size, negatively correlated with

projective-size. It is argued that size-constancy represents the basic functional relationship between object-size and perceived-size, in preference to the view that perceived-size is a compromise between object-size and angular retinal-image size.

References

BRUNSWIK, E. Die Zugänglichkeit von Gegenständen für die Wahrnehmung und deren quantitative Bestimmung. *Arch. ges. Psychol.*, 1933, **88**, 377–418.

BRUNSWIK, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.

CARLSON, V. R. Overestimation in size-constancy judgments. *Amer. J. Psychol.*, 1960, **73**, 199–213.

CARLSON, V. R. Underestimation in size-constancy judgments. *Amer. J. Psychol.*, 1962, in press.

GIBSON, J. J. The visual field and the visual world: A reply to Professor Boring. *Psychol. Rev.*, 1952, **59**, 149–151.

GILINSKY, A. S. The effect of attitude upon the perception of size. *Amer. J. Psychol.*, 1955, **68**, 173–192.

HOLADAY, B. E. Die Grössenkonstanz der Sehdinge bei Variation der inneren und äusseren Wahrnehmungsbedingungen. *Arch. ges. Psychol.*, 1933, **88**, 419–486.

HOLWAY, A. H., & BORING, E. G. Determinants of apparent visual size with distance variant. *Amer. J. Psychol.*, 1941, **54**, 21–37.

JENKIN, N. Effects of varied distance on short range size judgments. *J. exp. Psychol.*, 1957, **54**, 327–331.

JOYNSON, R. B. The problem of size and distance. *Quart. J. exp. Psychol.*, 1949, **1**, 119–135.

OSGOOD, C. E. *Method and theory in experimental psychology.* New York: Oxford Univer. Press, 1953.

SINGER, J. L. Personal and environmental determinants of perception in a size constancy experiment. *J. exp. Psychol.*, 1952, **43**, 420–427.

SMITH, W. M. A methodological study of size-distance perception. *J. Psychol.*, 1953, **35**, 143–153.

THOULESS, R. H. Phenomenal regression to the real object. *Brit. J. Psychol.*, 1931, **21**, 339–359.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.

(Received October 24, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 74–83

THE OPERATION OF SET IN A VISUAL SEARCH TASK [1]

HARRIET FOSTER

Engineering Psychology Laboratory, Institute of Science and Technology, University of Michigan

The purpose of the present study was to investigate the effects of "mental sets" on the speed of detection or identification of critical components of a visual display. Different types of sets with varying levels of specificity were investigated in order to determine their effects on visual search time and in the hope of gaining some understanding concerning the nature of set. A set is treated here as synonomous with an hypothesis, an expectancy of certain objects, features, or events. Set implies a restriction of the number of possible events for which one is "prepared." A particular set may be described or defined by the unifying concept or the category of objects, characteristics, or events for which the individual is prepared.

Two types of set were investigated in the present study: (*a*) functional sets, in which the categories were based on the meaning or significance of category members without regard to physical appearance, and (*b*) structural sets, in which the categories were based on the physical appearance of members. These two types of set are comparable to the functional and formal categories described by Bruner, Goodnow, and Austin (1956).

Bruner (1957) has suggested that set is a gating or filtering process whereby an individual is sensitive to certain stimulus features or certain events and relatively insensitive to others. One might expect that the efficiency of any particular set would depend upon the number of "gates" included in the set. It was hypothesized in the present study that functional categories require a "gate" for each category member, but that the gates for structural categories consist of the defining attributes or the observable common features of the category members.

With a functional set, then, the time required to recognize an individual category member is expected to increase as the number of category members increases, or as the category becomes broader, more general, or more inclusive. There are a good number of studies which support this hypothesis. Postman and Bruner (1949) and Freeman and Engler (1955) compared a double set for "food" and "color" words with a single set for color words. The double set, of course, contained more alternatives than the single set. Postman and Bruner found a higher mean recognition threshold for the multiple set; but Freeman and Engler found no significant difference between the multiple and single set. Rees and Israel (1935) compared two single sets with a double set which contained fewer alternatives than either single set and found the double set to be more effective than either of the single sets. The prediction concerning functional sets is also consistent with the numerous studies relating reaction time

[1] This work was conducted by Project MICHIGAN under Department of the Army Contract DA-36-039 SC-78801, administered by the United States Army Signal Corps. Particular thanks are due Sally Wurtz for her preparation of the word lists used in this study. The writer is also indebted to William L. Hays for his comments and suggestions concerning the statistical analysis of the data. Subjects in the pilot study were run by Sally Wurtz and in the main experiment by John T. Burns.

or duration threshold to stimulus or response uncertainty (e.g., Bricker, 1955; Hyman, 1953; Hyman & Hake, 1954; Krulee, Podell, & Ronco, 1954; Krulee & Weisz, 1955) and studies showing decreases in accuracy with increasing numbers of alternatives (e.g., Bruner, Miller, & Zimmerman, 1955; Long, Henneman, & Reid, 1953; Long & Lee, 1953a, 1953b; Long, Reid, & Garvey, 1954.)

Structural sets, on the other hand, are expected to result in increasing recognition times as the number of common attributes of category members increases. In terms of the generality of the category this means an increase in recognition time as the category becomes less inclusive or more specific. Binder's (1955) statistical model for visual recognition makes a similar prediction, but it is difficult to find much supporting empirical evidence. Some indirect support can be found in Chapman's (1932) results showing a loss on attributes for which *S* is not set, and in the results of Lawrence and Laberge (1956) which showed a slight (but not significant) advantage for a set for only one dimension over a set for all dimensions of a stimulus. More direct support can be found in a study by Archer (1954) in which *Ss* were required to classify each of a series of visual patterns on four dimensions. The time required to classify the patterns was found to increase with the number of relevant stimulus dimensions and at the same time did not change significantly with different amounts of irrelevant information. On the other hand, evidence summarized by Miller's (1956) "chunking" analysis suggests that a complex stimulus may be responded to as a unit. In this case, no differences in recognition time would be expected as a result of varying numbers of defining attributes of sets or categories.

In addition to the set variables two recognition tasks were studied. Subjects were asked to detect or locate the category member or they were required to identify or specify the particular category member which was present. It was hypothesized that with functional sets, detection would be equivalent to identification. If the gates of a functional set represent individual category members, then the matching of incoming information with one gate is equivalent to the identification of the object. On the other hand, it was expected that detection would be faster than identification when a structural set was used, and that the difference would be greatest for the broadest set or the smallest number of defining attributes. Here, it is supposed, the gate ignores the unique features of one member—the features which must be used to distinguish or identify the particular category member.

Method

Stimulus materials.—Printed words were used as the stimulus materials. The functional characteristics (meanings) of printed words are well-known to *Ss* as are their structural characteristics (taken here to be the letters which make up the words). The relative inclusiveness of functional categories was varied by selecting general categories and subcategories of the more general categories, whereas the inclusiveness of structural categories varied with the number of letters included in the set. Printed words may represent the most extreme case of a lack of correspondence between structure and function and because of this they may be the best medium for studying the ability of *Ss* to respond to functional characteristics of visual targets.

Six functional categories, each with two subcategories, were chosen and lists of four-, five-, and six-letter words falling within these categories were compiled. The Thorndike-Lorge (1944) general word count was used to equate word familiarity of two sets of three lists each (Lists A-1, B-1, and C-1, and Lists A-2, B-2, and C-2). A list compiled by Cohen, Bousfield, and Whitmarsh (1957) was used in an attempt to equate the degree of association of the words with their respective categories, but there were not sufficient empirical data for complete control of this variable.

Each word was presented on a different slide along with eight unrelated words of equal length in a 3 × 3 display, marked off by grid lines into nine cells. The unrelated or background words varied in familiarity (Thorndike-Lorge count), with approximately the same distribution of background word frequency on each slide. No one word appeared on more than one slide. The positions of critical and background words were randomly assigned to the nine locations with the restriction that critical words within any one category list would occur in each of nine positions at least once and not more than twice. All letters of words were capitalized.

Response measures.—Because the number of response alternatives will vary with the broadness of the category or between detection and identification, an attempt was made to control response choice time by allowing *S*s but one response—that of pressing a hand switch held in the left hand. The time measured was the time required by *S* to indicate that he "knew" the correct answer but not the time required to make the correct response. It was hoped that in this way differences between experimental groups would be associated primarily with "perceptual" rather than with response speed. A timer was started when the slide came on; the timer stopped and the slide went off when *S* pressed his hand button.

The *S*s were divided into two response groups—the Detection group (Group D) and the Identification group (Group I). Detection *S*s were instructed to indicate the location of the correct word. As soon as *S* knew the location he pressed his hand switch, which turned off the timer and slide. He then indicated the location of the correct word by pressing with the right hand one of nine buttons on a panel in front of him. The nine buttons were arranged in a 3 × 3 matrix corresponding to the arrangement of words in the slide.

Identification *S*s were instructed to tell *E* what the correct word on the slide was. These *S*s pressed the hand switch as soon as they knew what the word was, and then said the word aloud to *E* after the timer had stopped.

Procedure.—The experimental design is summarized in Table 1. Groups D and I were each subdivided into three set groups. The different lists or categories were presented in the same order for all subgroups. One order of presentation within lists or categories was used for half of the group and the reverse order for the other half of the group. The order in which the different sets were adopted

TABLE 1

EXPERIMENTAL DESIGN

Group	Category or List								
	Functional Categories						Structural Categories		
	A-1 (Animals) 15 Words	A-2 (Transportation) 14 Words	B-1 (First Names) 15 Words	B-2 (Plants) 14 Words	C-1 (Geographical) 15 Words	C-2 (Food) 14 Words	1 9 Words	2 9 Words	3 9 Words
Detection 1_d	Specific Set	Specific Set	Broad Set	Broad Set	Narrow Set	Narrow Set	Very Narrow Set	Broad Set	Narrow Set
Detection 2_d	Narrow Set	Narrow Set	Specific Set	Specific Set	Broad Set	Broad Set	Narrow Set	Very Narrow Set	Broad Set
Detection 3_d	Broad Set	Broad Set	Narrow Set	Narrow Set	Specific Set	Specific Set	Broad Set	Narrow Set	Very Narrow Set
Identification 1_i	Double Set	Double Set	Broad Set	Broad Set	Narrow Set	Narrow Set	Very Narrow Set	Broad Set	Narrow Set
Identification 2_i	Narrow Set	Narrow Set	Double Set	Double Set	Broad Set	Broad Set	Narrow Set	Very Narrow Set	Broad Set
Identification 3_i	Broad Set	Broad Set	Narrow Set	Narrow Set	Double Set	Double Set	Broad Set	Narrow Set	Very Narrow Set

Note.—A warm-up and posttest task involving four-digit numbers preceded and followed the main experiment (see text).

and the particular set which was used for any one list or category differed for the different subgroups. Three types of set were used for the functional categories in Group D: (*a*) a specific set in which *S* was told the exact word for which he was to look, *E* pronouncing and spelling each word; (*b*) a narrow set in which *S* was told the subcategory to which the word belonged (e.g., in Category A-1, members of the animal kingdom, two subcategories were used—four-legged animals and birds); and (*c*) a broad set which included the two subcategories as well as other subcategories (e.g., the animal category included fish as well as four-legged animals and birds). Each narrow or broad category was defined for *S* and then illustrated by three (narrow categories) or four (broad categories) examples. The sets for structural categories were either broad (the first letter specified), narrow (the first two letters specified), or very narrow (the first three letters specified).

The sets prescribed for the subgroups of Group I were the same as Group D, with the exception of the specific set. Because Group I was instructed to specify the exact word which appeared, a specific set instruction would have been meaningless; *S* would have been told exactly what he was to say. In place of a specific set, subjects of Group I were given a double set. They were instructed to look for words which fell into either of the two subcategories. For example, for Category A-1, *S*s were told to look for a word which was *either* a four-legged animal *or* a bird. The sets for structural categories were the same as Group D.

Subgroups differed only in the sets which were adopted. The critical words, their positions on the slides, and their background words were identical for all groups, with the following exceptions: When *S*s were given a broad functional category an additional five slides were introduced to insure that *S*s included in their sets additional subcategories. When *S*s used a narrow or a very narrow structural set, two (narrow set) or one (very narrow set) extra slides were introduced which contained a word in the background which began with the same letter or same two letters as the critical word but which did not contain all of the critical letters. The responses to these extra slides are not included in the data analysis.

The sets assigned for the different categories and groups are indicated in Table 1.

All groups received 27 trials on a warm-up task and 27 trials on a posttest. The stimuli on these two tasks were four-digit numbers arranged in a 3 × 3 matrix as in the case of the words. On each trial *S* was told the first two digits of the correct number. The *S*s of Group D were required to indicate the location of the correct number after they had turned the slide off by means of the hand switch. The *S*s of Group I said the entire number aloud after turning off the slide with the hand switch. The major purpose of the warm-up trials was to reduce variability due to practice in the main experimental trials. The posttest was introduced in an attempt to get some measure of practice or fatigue effects from the beginning to the end of the experimental session. The two tests may also give some indication of individual differences in reaction times within and between groups.

Subjects.—The *S*s were 72 University of Michigan students who had answered an advertisement in the college paper. They were paid according to the amount of time spent and the quality of their performance. As much as was possible, *S*s were assigned successively to the different subgroups, so that 1 *S* was run in each group before a second *S* was run in any group. The subgroups were composed of 12 *S*s each, 6 men and 6 women, with the exception of Group I-2 and Group I-3 and Group D-1, each of which contained 7 men and 5 women.

Apparatus.—The *S* was seated at a console which contained a rear projection screen 8 in. high and 11.5 in. wide. A sloping table projected 10.5 in. from below the projection screen. In the center of this table was the panel containing the 3 × 3 matrix of buttons which *S*s in Group D used to indicate the location of the critical word or number. This response panel was covered by a black box when *S*s in Group I were run. The *S*s held in their left hands the switch which was easily pressed with one finger. They sat in a swivel chair, the legs of which were in a fixed position. The height of the chair was adjusted for different *S*s so that their eyes would be approximately in line with the center of the projection screen. The *S*s wore earphones during the experiment, through which white noise was fed except when *E* was giving instructions. Instructions were also given through the earphones.

The *E*'s panel on the opposite side of the console consisted of a 35-mm. slide projector with magazine loading, a time clock calibrated in .01 sec., a row of nine numbered lights which indicated which of the nine location buttons had been pressed by *S*, and a microphone and foot switch which were used for giving instructions and turning off the white noise.

TABLE 2

MEANS OF LOG MEDIAN RESPONSE TIMES FOR CORRECT RESPONSES ON FUNCTIONAL CATEGORIES

Set	List												Set Mean
	A-1		A-2		B-1		B-2		C-1		C-2		
	Group	Mean	Group	Mean	Group	Mean	Group	Mean	Group	Mean	Group	Mean	
Detection Task													
Broad	3	2.269	3	2.306	1	2.272	1	2.317	2	2.369	2	2.302	2.306
Narrow	2	2.286	2	2.325	3	2.216	3	2.258	1	2.331	1	2.283	2.283
Specific	1	2.114	1	2.073	2	2.116	2	2.094	3	2.087	3	2.086	2.095
List mean		2.223		2.235		2.201		2.223		2.262		2.224	
Identification Task													
Broad	3	2.290	3	2.320	1	2.223	1	2.286	2	2.367	2	2.276	2.293
Narrow	2	2.277	2	2.343	3	2.202	3	2.301	1	2.316	1	2.320	2.278
Double	1	2.289	1	2.302	2	2.232	2	2.286	3	2.327	3	2.285	2.295
List mean		2.285		2.322		2.219		2.291		2.337		2.264	

Note.—Group means for detection task were: 1 = 2.232, 2 = 2.212, 3 = 2.204; group means for identification task were: 1 = 2.274, 2 = 2.305, 3 = 2.287.

Instructions.—Before the warm-up trials *S* was given instructions either to locate or to identify the correct stimulus. He was told that his score would depend upon how quickly he made the correct response and that he would get no credit for incorrect responses. Before each category, *S* was instructed as to what the category was. In the case of the specific set or the structural sets he was told the word or letter(s) preceding the presentation of each slide. For the specific set the word was pronounced and spelled.

RESULTS

Median times for correct responses, in .01 sec., were calculated for each *S* for each of the nine category lists. In order to test the normality of the distribution of medians, each median was transformed to a standard score for the particular task, list, and set in which it occurred. The standard scores were then combined into one distribution and tested for normality by means of a χ^2 analysis. Median times resulted in a distribution of scores significantly different from normal ($P = .01$). Log transformations of median scores, tested in the same manner, did not differ significantly from a normal distribution ($P > .10$). All analyses were therefore made with the log transformed scores.

Functional categories.—Mean scores for the functional categories are shown in Table 2. Warm-up and posttest scores appear in Table 3. Basically the experimental design consists of four latin squares, one for each task and one for each type of category set. However, a latin square analysis for the functional sets would require that pairs of lists be combined. Because the two lists of each pair differed in word frequencies the lists

TABLE 3

MEANS OF LOG MEDIAN RESPONSE TIMES ON WARM-UP AND POSTTEST

Group	Warm-up		Posttest	
	Mean	*SD*	Mean	*SD*
D-1	2.191	.067	2.182	.064
D-2	2.210	.051	2.174	.060
D-3	2.179	.048	2.177	.065
I-1	2.199	.034	2.189	.038
I-2	2.208	.069	2.206	.064
I-3	2.202	.082	2.186	.055

were not combined, but, instead, a cross-over analysis was used (Cochran & Cox, 1957). In this way it was possible in one analysis to treat the six lists separately. Table 4 presents the results of the analysis for Groups D and I. Variations in broadness of set result in significantly different response times in Group D but not in Group I. It should be recalled, however, that Group D, used a specific set when Group I used a double set. The significance of the results in Group D appears to be due primarily to the difference between the specific set and the two broader sets. The mean across the six lists is higher for the broad set than for the narrow set, but the difference is small. The same direction of difference between broad and narrow sets is found in the performance of Group I. In both cases there are reversals in the direction of the difference on two of the individual lists. Group differences may be partially responsible for these reversals.

Although the double set results in a slower response time on the average than the narrow set, the difference is not significant, nor is the difference consistent for all lists.

The number of errors made with functional sets was small but showed the same order of difficulty of sets. The number of possible errors with all words, lists, and *S*s combined is 1044 for each task and set. Detection *S*s made 17 errors with the broad set, 9 errors with the narrow set, and 4 errors with the specific set. Identification *S*s made 19 errors with the broad set, 15 errors with the narrow set, and 23 errors with the double set.

Examination of the size of the differences between sets on different lists suggests a possible interaction effect between sets and lists. The experiment was not designed to test interaction effects, and because of the lack of independence between cells a really adequate test can not be made. An attempt was made, however, to discover the presence of an interaction. Each of the three groups on each task was subdivided into six groups of 2 *S*s each. Each pair of *S*s contained 1 *S* from each of the two slide presentation orders. Otherwise the *S*s were randomly paired. Each

TABLE 4

EFFECTS OF VARIATIONS IN BROADNESS OF SET

Source	Functional Categories			Structural Categories		
	df	Group D *F*	Group I *F*	*df*	Group D *F*	Group I *F*
Set	2	127.37**	1.64	2	3.45	35.26*
Lists	5	1.90	20.79**	2	.56	4.12
Groups	2	4.91*	5.28*	2	4.50	59.19*
Error	8	(.000631)[a]	(.000266)[a]	2	(.000327)[a]	(.0000215)[a]

[a] Error mean squares.
* $P < .05$.
** $P < .01$.

pair was randomly assigned to one of the six list and set combinations of its group. Such a procedure, of course, reduces the reliability of the results to a marked degree, as the data from any 1 *S* is being used for only one of the six lists. However, it is possible in this way to obtain independence of the cells and examine the data for interaction effects. The analysis of variance involves a double classification, with three variations of set and six different category lists. Groups D and I were analyzed separately due to the difference in set for the specific and double categories. The analysis for Group D again showed a significant difference between sets, but no significant list variation was found for Group I. The results of the analyses failed to show significant interactions between lists and sets. However, because the number of observations included in the analyses was so greatly reduced, the possibility of an interaction effect remains.

Task differences were not expected on the functional categories. Four separate analyses of variance, required by the lack of independence between cells, were made to test differences between Groups D and I on three lists. Separate analyses were made for the broad and narrow sets and the two sets of matched-frequency word lists. The *F* values for the four analyses are not, of course, independent of each other, as the same *S*s are being compared in each. The four *F* values for the task variables were all less than 1. The *F* values for lists were significant at the .01 level in three of the analyses but were not significant for Lists A-2, B-2, and C-2 when the broad set was used. There were no significant interactions between lists and tasks.

Structural categories.—Mean log times for structural categories are shown in Table 5. The latin square analyses for Groups D and I are presented in Table 4. Group I shows an increasing response time as the structural category becomes more

TABLE 5

MEANS OF LOG MEDIAN RESPONSE TIMES FOR CORRECT RESPONSES ON STRUCTURAL CATEGORIES

Set	List						Set Mean
	1		2		3		
	Group	Mean	Group	Mean	Group	Mean	
Detection Task							
Broad	3	2.127	1	2.155	2	2.162	2.148
Narrow	2	2.193	3	2.150	1	2.217	2.187
Very narrow	1	2.164	2	2.194	3	2.151	2.170
List mean		2.161		2.166		2.177	
Identification Task							
Broad	3	2.120	1	2.159	2	2.167	2.149
Narrow	2	2.184	3	2.148	1	2.175	2.169
Very narrow	1	2.175	2	2.202	3	2.163	2.180
List mean		2.160		2.170		2.168	

Note.—Group means for detection task were: 1 = 2.179, 2 = 2.183, 3 = 2.143; for identification task: 1 = 2.170, 2 = 2.184, 3 = 2.144.

specific, or as the number of defining attributes increases. The effect of variation in broadness of set is significant ($P = .05$). Although Group D tends to respond more slowly with the narrow set than with the broad set, times for the very narrow set are somewhat faster than those for the narrow set. The differences in broadness of set are not significant for Group D. The total number of errors for the different sets and tasks is too small to have any significance. The largest number of errors is a total of 3 out of a possible 324 for the narrow set in each of the recognition tasks.

It was predicted that Group I would be slower than Group D when structural sets were used. However, analyses of variance, computed separately for each of the three sets, yielded task *F*s all less than 1. Interactions between sets and lists were tested in the same manner as for the functional categories, but because there were only three lists instead of six the task groups were subdivided into three groups of 4 *S*s each. And because all three set variations were comparable in Groups D and I, it was possible to use a triple classification of tasks, lists, and broadness of set. None of the *F*s was significant. The reader is reminded, however, that the number of observations was reduced to one-third of the original number in order to obtain independent cell scores.

Comparison of structural and functional categories.—The present experiment was designed to measure the effects of variations in the broadness of set within structural and functional categories and not to provide a direct comparison between structural and functional categories. The words used for the two types of category differed, and no direct comparison can be made. However, a small number of *S*s was run in a pilot study in which each of 31 words was tested under each of three sets—broad functional, narrow functional, and broad structural. Functional sets resulted in times significantly longer than times for structural sets.

Discussion

It was hypothesized that the efficiency of set in a visual search task would be a function of the number of "gates" which were required by the set. The gates for functional categories were expected to be the individual category members, leading to increasing recognition times as the category became broader or more inclusive. Variation in structural set was expected to vary with the number of defining attributes of the category, and response times were expected to increase as the category became more specific. The results of this experiment tend to support the hypotheses, but to a limited degree.

Differences between broad and narrow functional sets were small. Although it is not possible to measure the difference in information value between the broad and narrow sets, it seems safe to assume that the difference is much smaller than the difference in information value bebetween narrow and specific sets. A more adequate test of the broad-narrow set difference might be made by selecting a narrow category with fewer members, thus increasing the difference in the number of alternatives of broad and narrow categories. For example, a narrow category under the general category of animals might be "mammals" or "members of the cat family." One would expect the effect of differences in set to be more pronounced in such a comparison.

It is possible that the failure to find significant differences for functional categories in the identification task is a function of the nature of the task, rather than of the small differences in broadness of set. Such a possibility seems remote, however, when one considers the lack of difference between detection and identification performance.

The results of this study give little indication of the direction of difference

between broad and double sets. If the number of category members is the only determinant, then the broad set should be less efficient. However, it is possible that the disjunctive character of the double set introduces additional difficulties not found with a single set.

It was originally hypothesized that the use of functional categories would result in separate gates for each category member, whereas the different gates for structural categories would be the defining attributes of the categories. The hypothesis concerning structural categories however, is contrary to the fairly large body of literature summarized in Miller's (1956) chunking analysis. Furthermore, the predictions concerning functional and structural sets are inconsistent with each other for very specific categories. A completely specified word if it is considered as a one-member functional category requires the use of only one gate, whereas if it is treated as a specific structural category it requires a gate for each letter. Although the detection times for the very narrow structural set are not significantly shorter than those for the narrow structural set, the results suggest that a switching point may be reached in structural sets. If either type of gate can be used, and the results of this study indicate that they can, then the most efficient procedure with structural sets would be to gate in terms of attributes until the number of attributes becomes larger than the number of category members (supposing that the two types of gates are in a sense equivalent) at which point the gating should then be done in terms of category members.

No significant difference was found in this study between the detection and identification tasks, either with functional or structural categories. The identification required in this task was much more simple than is often the case. All of the cues necessary for accurate identification were present, easy to discriminate, and unambiguous. Perhaps a more difficult identification would have resulted in differences, although there is nothing in the data presented here to indicate that this would be so.

Summary

The study was designed to determine the effects of varying sets on the location and identification of words in a field of unrelated words. Sets were defined in terms of either functional or structural categories with varying degrees of specificity.

In general, response times to functional categories decreased as the categories became more specific, or as the number of category members decreased. Differences between specific word definitions and the broader category definitions were pronounced. Differences between broad general categories and their narrower subcategories were small, but in the predicted direction. Words defined by structural categories tended to be identified more rapidly when the category contained the smallest number of defining attributes. Some ambiguity was found with greater numbers of defining attributes.

The results are interpreted in terms of a "gating" mechanism, where the "gates" for functional categories are the category members and the gates for structural categories are the defining attributes of the category.

References

Archer, E. J. Identification of visual patterns as a function of informational load. *J. exp. Psychol.*, 1954, **48**, 313–317.

Binder, A. A statistical model for the process of visual recognition. *Psychol. Rev.*, 1955, **62**, 119–129.

Bricker, P. D. The identification of redundant stimulus patterns. *J. exp. Psychol.*, 1955, **49**, 73–81.

Bruner, J. S. On perceptual readiness. *Psychol. Rev.*, 1957, **64**, 123–152.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.

Bruner, J. S., Miller, G. A., & Zimmerman, C. Discriminative skill and discriminative matching in perceptual recognition. *J. exp. Psychol.*, 1955, **49**, 187–192.

Chapman, D. W. Relative effects of determinate and indeterminate Aufgaben. *Amer. J. Psychol.*, 1932, **44**, 163–174.

Cochran, W. G., & Cox, G. M. *Experimental designs.* (2nd ed.) New York: Wiley, 1957.

Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. *Cultural norms for verbal*

items in 43 categories. Technical Report No. 22, August 1957, University of Connecticut, ONR Contract Nonr-631(00).

FREEMAN, J. T., & ENGLER, J. Perceptual recognition thresholds as a function of multiple and single set and frequency of usage of the stimulus material. *Percept. mot. Skills*, 1955, **5**, 149–154.

HYMAN, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

HYMAN, R., & HAKE, H. W. Form recognition as a function of the number of forms which can be presented for recognition. *USAF WADC tech. Rep.*, May 1954, 54–164.

KRULEE, G. K., PODELL, J. E., & RONCO, P. G. Effect of number of alternatives and set on the visual discrimination of numerals. *J. exp. Psychol.*, 1954, **48**, 75–80.

KRULEE, G. K., & WEISZ, A. Studies in the visual discrimination of multiple-unit displays. *J. exp. Psychol.*, 1955, **50**, 316–324.

LAWRENCE, D. H., & LABERGE, D. L. Relationship between recognition accuracy and order of reporting dimensions. *J. exp. Psychol.*, 1956, **51**, 12–18.

LONG, E. R., HENNEMAN, R. H., & REID, L. S. Theoretical considerations and exploratory investigation of "set" as response restriction. *USAF WADC tech. Rep.*, December 1953, 53–311.

LONG, E. R., & LEE, W. A. The influence of specific stimulus cuing on location responses. *USAF WADC tech. Rep.*, December 1953, 53–314. (a)

LONG, E. R., & LEE, W. A. The role of spatial cuing as a response-limiter for location responses. *USAF WADC tech. Rep.*, December 1953, 53–312. (b)

LONG, E. R., REID, L. S., & GARVEY, W. D. The role of stimulus ambiguity and degree of response restriction in the recognition of distorted letter patterns. *USAF WADC tech. Rep.*, May 1954, 54–147.

MILLER, G. A. The magical number seven, plus or minus two: Some limitations on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

POSTMAN, L., & BRUNER, J. S. Multiplicity of set as a determinant of perceptual behavior. *J. exp. Psychol.*, 1949, **39**, 369–377.

REES, J. J., & ISRAEL, H. E. An investigation of the establishment and operation of mental sets. *Psychol. Monogr.*, 1935, **46**(6, Whole No. 210).

THORNDIKE, E. L., & LORGE, L. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

(Received October 24, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 84–90

ANTICIPATORY TIMING OF CONTINUOUS AND DISCRETE RESPONSES[1]

JACK A. ADAMS AND LYLE R. CREAMER

University of Illinois

Helson (1949, p. 496) appears to have been the first to observe that response to a repetitive input signal in continuous visual tracking tasks has timing errors that are less than would be expected on the basis of visual-motor reaction time (RT). With practice, *S* learns to anticipate regular changes in the input and he achieves a higher performance level than when anticipation is absent and response awaits the visual detection of a change, thus producing tracking error proportional to RT. This type of learned anticipation in paced tasks without preview is what Poulton (1952a, 1952b, 1957) calls *perceptual anticipation*. Adams and Xhignesse (1960) make the additional distinction of *beneficial anticipation*, and define it in a measure of learned response timing that is associated with high skill and is distinct from premature anticipations that can produce performance poorer than if *S* merely had responded after the occurrence of a signal change as in classical RT. Bartlett (1951), Vince (1953, 1955), and Adams (1961) provide further discussions of anticipation in skilled performance.

There is nothing but unverified hypotheses on the underlying mechanisms of perceptual anticipation. Vince (1953) discusses the cognitive basis of these learned anticipatory mechanisms. Adams (1961) discusses as one possibility an S-R hypothesis where an anticipatory overt tracking response is determined by the action of a mediating response and its response-produced cues, which is an explanation similar to that advanced for the effects of verbal pretraining (Arnoult, 1957; Goss, 1955; Vanderplas, 1958). If this hypothesis deserves serious consideration, we should be able to demonstrate that pretraining in anticipation with a nontracking response should show positive transfer to measures of perceptual anticipation for the tracking response in the criterion task. Our experiments reported here were designed to go beyond a demonstration of transfer by further asking whether mediating responses are central-verbal or peripheral-motor responses. Of course the mediation view could be wrong, and anticipation could depend on time-varying concomitants of the criterion motor response itself, and not pretrainable as mediational views would require. If so, we would expect no differences for our various pretraining activities.

[1] This research was supported by the United States Air Force under Contract No. AF 49(638)-371, monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

EXPERIMENT I

Method

The basic task was one-dimensional continuous pursuit tracking, with the criterion measure being *S*'s anticipation of directional change in a regular input signal which approximated a sine wave. Experimental groups were given special pretraining in anticipation of directional change and then were transferred to the whole task. Comparisons on the criterion response were with a control group having only whole-task practice.

Apparatus.—The signal to be tracked was a cam-generated red line on moving white paper which appeared in a narrow horizontal

window below eye level measuring 7 mm. by 30 cm. Wave frequency was 34 cycles per min., and the scale of the signal gave it a lateral sweep of 20 cm. The standing *S* responded with a horizontal handle having a radius of 16.6 cm., and the motor movement was a back and forth arc at about the level of an ordinary table. There was a mechanical linkage between the control and a blue pen writing on the moving paper, and tracking error at any moment was the discrepancy between the red and blue lines. The *E* used two Standard Electric timers for controlling the duration of trials and intertrial rests.

For pretraining in anticipation, auxiliary response apparatus included a throat microphone for verbal responding, a button of 2 cm. diameter mounted next to the control hand for discrete motor responses, and a small crank with a radius of 7.5 cm. for continuous responding. The particular application of each of these will be explained below, but for all of these conditions we followed the procedure of deceiving *S* to think that his pretraining responses were being recorded. This ego-involving approach was used to minimize desultory responding which did not involve active attempts to anticipate directional changes of the signal. Toward this end, each of these pretraining response devices had obvious wires leading from them and out a hole in the wall, and *S* was told that his performance was being measured. These methods were patterned after those of Adams (1955) where similar deceptions were used successfully in guaranteeing *S*'s active participation.

Procedure.—Each trial was 60 sec. in duration and intertrial rest was 30 sec. The three experimental groups had 12 pretraining trials, a 12-min. rest, and 10 whole-task criterion trials. The control group (WT) had only the 10 whole-task trials.

The following is a description of the three different pretraining activities:

Group V (verbal).—The *S* responded by saying the word "Change" into the throat microphone each time the target line reversed direction. He was instructed to anticipate the change so his verbal response coincided with the exact instant of the directional reversal of the signal. The purpose of this group was to see if the mediators were verbal-central and independent of the arm-hand effector used in tracking.

Group CM (continuous motor).—An *S* had a small crank set in a plane perpendicular to the larger control handle of the criterion task. The task was to match the speed and direction change of the input signal, and *S*'s response was a back and forth movement of the small crank over a span of about 180°. Gibbs (1954a, 1954b) has discussed the continuous functions for neural discharge when a voluntary movement is made, and it is plausible that the mediating mechanism for anticipation depends on some aspect of this continuous neural output of the primary motor effector system used in tracking. A pretraining series, where *S* attempts to duplicate the rate and direction of the signal, might prove to be favorable conditions for the acquisition and transfer of anticipatory behavior.

Group DM (discrete motor).—The *S* pressed a button each time the target reversed direction. Like Group V, they were urged to anticipate the directional change so that their press of the button coincided exactly with the change. This was a control for Group CM to cover the possibility that any response with the motor effectors of the transfer tracking task is a sufficient condition for acquiring anticipatory behavior.

Subjects.—The *S*s were 47 university male undergraduates, and they were paid for their participation. The *S*s were randomly assigned to groups. Each group had 12 *S*s except Group DM which had 11. One *S* was discarded because of failure to understand the instructions.

Results

The method of measurement was the same as used by Poulton (1952b) and Adams and Xhignesse (1960). The response time measure was taken as the time in seconds between a directional change of the input signal and a corresponding change in direction of *S*'s response movement as he attempts to follow the reversal. We followed the algebraic convention of a positive sign when response preceded the signal change (lead error) and a negative sign when it followed the change (lag error). A zero value was perfect anticipation. Starting with the third cycle on a trial, measures were taken for the next 60 responses to directional change of the signal.

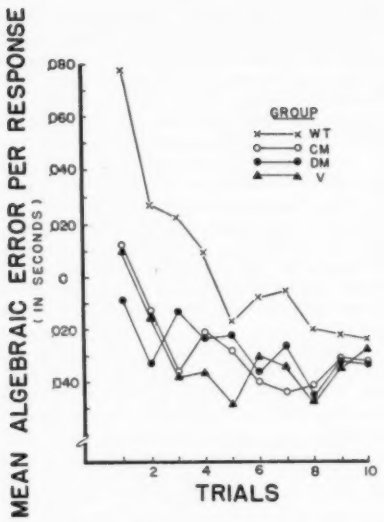

FIG. 1. Performance on the continuous pursuit tracking apparatus used as the transfer task in Exp. I.

Figure 1 is a plot of mean algebraic error for the 10 whole-task trials. The pretraining activities had a substantial positive transfer effect on whole-task performance because the three experimental groups all have less mean error throughout than the Group WT. Using total algebraic error for the 60 responses as *S*'s score on Trial 1, simple analysis of variance for all groups gave an F ratio of 4.45 ($df = 3/43$, $P < .01$). Tukey's gap test (Snedecor, 1956, p. 251) showed that the experimental groups did not differ significantly among themselves (.05).

The effect of the pretraining was to reduce the lag error which is so prominent for Group WT on Trial 1. Of course the mean algebraic error in the vicinity of zero for the three experimental groups could come from an appropriate distribution of relatively large plus and minus values without any beneficial anticipation whatsoever. To check this, we defined beneficial anticipation as a response to directional change occurring in the range of ± 76 msec. This range was chosen because the smallest unit of measurement on our recording paper was 1 mm. which, at the chart speed we used, was 76 msec. Responses whose values were -76, 0, or $+76$ msec. were categorized as beneficially anticipatory because they are clearly less than any known human RT value. Our next level of response categorization would have been ± 2 mm. or ± 152 msec. but, since RT under some conditions can be less than 152 msec. we would not have been so confident of an index of beneficial anticipation had this gross criterion been used. The percentages of beneficial anticipations, based on the total number of responses made by each group on Trial 1, were greater for pretraining groups. The percentages for Groups WT, V, DM, and CM were 38, 52, 55, and 55, respectively. For statistical testing, a beneficial anticipation score was defined for each *S* as the number of beneficial anticipations on Trial 1. These scores for the four groups were significantly different at the .05 level ($F = 3.41$, $df = 3/43$). The gap test found no difference among Groups V, DM, and CM. We conclude that the Trial 1 differences for algebraic error in Fig. 1 are due mainly to beneficial anticipatory responses acquired in pretraining, but that pretraining did not function to differentiate among the three experimental groups.

EXPERIMENT II

Discrete tracking differs from continuous tracking in that each event is in a stable state. Anticipation of event change is based solely on time estimation, in contrast to continuous tracking where time estimation and changes in the signal pattern both

occur as cues for anticipatory responses. Experiments II and III were run to broaden the scope of our findings in Exp. I. The results of Exp. II were negative and will be discussed succinctly as background for the positive findings of Exp. III.

Method

The Discrete Tracking Apparatus (DTA) was used, and is described under Exp. III. The stimulus sequence was a repetition of three events. The number of pretraining and whole-task trials, as well as trial duration and intertrial rests were the same as Exp. I. A control group (WT) had just whole-task visual discrete tracking. Four experimental groups in the pretraining trials either made a verbal response to predict the onset of each of three visual signals (VV), a verbal response to each of three auditory signals (AV), a visual-motor response of pressing one of three buttons to coincide with the onset of each visual signal (VM), or the same button pressing motor responses to auditory signals (AM). The *S*s were told that their pretraining performance was being scored. Each group had 8 male *S*s.

Results

The left-hand side of Fig. 2 shows the results in terms of time on target for the whole-task trials. While time on target ordinarily is used as an overall measure of motor skill in a tracking task, in discrete tracking *S*'s proficiency is directly a function of timing an elementary positioning response to a stimulus change, with none of the elaborate response sequences required in continuous tracking. With a very simple repetitive stimulus sequence and with only rare wrong movements, as was the case here, off-target time is a matter of the accuracy of timing the response to each stimulus change, and so time on target becomes a direct index of anticipation. Perfect anticipation of each event would give 100% time on target. A simple analysis of variance test of time-on-target scores on Trial 1 lacked significance at the .05 level.

A beneficial anticipation score was

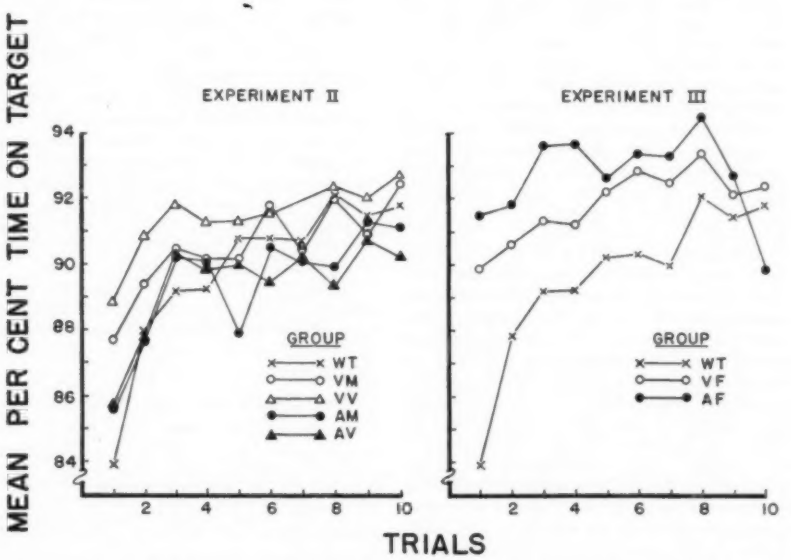

FIG. 2. Performance on the visual discrete tracking apparatus used as the transfer task in Exp. II and III.

determined for each *S* on Trial 1 of the whole-task, and a simple analysis of variance test of them for the five groups failed the .05 level. The chart units of the operations recorder used to record all stimulus and response events on Trial 1 were slightly different than in Exp. I and necessitated a criterion of ±67 rather than ±76 msec.

Experiment III

Method

We surmised that the negative findings of Exp. II were ascribable to poor knowledge of results. There was no feedback for any of the pretraining responses and it would be reasonable to expect that a mediating response as an explanation of perceptual anticipation would require knowledge of results for its learning. Experiment III was designed to see if anticipatory behavior is a function of feedback and, toward this end, some departure from the designs of Exp. I and II was used. Instead of having a neutral nontracking response for pretraining, pretraining in Exp. III was with a different version of the discrete tracking task so that feedback on the timing of responses could be had. The control was completely different in pretraining, but nevertheless allowing practice with feedback prior to transfer and practice in the criterion visual discrete tracking task with the stick control. Both of the controls were very simple, and very different, so we expected that amount of transfer would be determined by the time relationships that are learned in pretraining. Pretraining was with both auditory and visual stimuli to check the generality of learned time keeping with respect to sensory mode.

Apparatus.—The DTA in a two-dimensional visual version has been described elsewhere (Adams & Xhignesse, 1960), and the essential details of the earlier description apply for our use of it in a one-dimensional pursuit tracking configuration. Briefly, in this discrete tracking task, a stimulus appeared which was stable for a specified time, like a light or a tone, and *S* was required to position a control so that an error indication was nulled. The one-dimensional visual version of the DTA had a green, a white, and a red jeweled lamp of ½-in. diameter in a small display box in front of the seated *S*. One of these lights came on and *S* in the criterion task had to position a 2-in. control stick on the armrest of his chair so that a small feedback light beneath the cue light was on, signifying proper positioning of the control. If *S* was off target in one of the two wrong positions, the feedback light was under a cue light which was not illuminated (error cue) and he had to correct the error by moving his control to place the lighted feedback lamp under the cue light which was on. The cue light kept changing and the task of *S*s was to keep error to a minimum by having the feedback light in alignment with the cue light as much as possible.

In an auditory tracking version of this task three tones were presented sequentially over a headset, either 600, 800, or 1000 cycles. When *S* was on target he heard a pure tone, but when he was off target he heard a complex tone made up of a pure tone signifying the correct signal and, superimposed on it, one of the other two signals, rapidly interrupted, indicating the misalignment of the control.

The input series were regular repetitions of the three 1-sec. signals and were programed on punch tape and read by an automatic tape reader. Successive events were perfectly contiguous in time. This required a regular series of three discrete movements from *S*, back and forth, and beneficial anticipation was revealed in the closeness of the stimulus change and the response to it. Performance measurement was time on target, as well as complete recording of all stimulus events and responses with an Esterline-Angus operations recorder on Trial 1 of whole-task practice. Pretraining was with a rotary knob control having three distinct detent positions.

Procedure.—Number of trials, and the duration of trials and intertrial rest intervals, were the same as Exp. I and II. Group VF had visual tracking and feedback in the pretraining trials, and Group AF had auditory tracking and feedback in pretraining. Criterion whole-task practice for both groups was visual tracking with the stick control. Group WT from Exp. II was the control group.

Subjects.—The *S*s were 16 male university students. Eight were randomly assigned to each of the two groups. They were paid.

Results

The right-hand side of Fig. 2 has the results for time on target. In contrast to Exp. II, the pretraining with feedback gave a large amount of positive transfer, with Trial 1 performance for Groups VF and AF being about at Group WT's terminal performance level. A simple analysis of variance for time-on-target scores

on Trial 1 gave a significance level beyond .01 ($F = 13.04$, $df = 2/21$).

Using the same criterion for beneficial anticipation as Exp. II, the percentages of these precisely timed responses for Groups WT, VF, and AF on Trial 1 were 39, 67, and 73, respectively. A simple analysis of variance for beneficial anticipation scores gave an F ratio significant beyond the .01 level ($F = 11.09$, $df = 2/21$). The large positive transfer on Trial 1 was importantly a function of the relatively large number of beneficial anticipations. Tukey's gap test showed no significant difference between the two experimental groups.

Discussion

Positive transfer of pretraining to the accurate and anticipatory timing of directional changes in continuous visual tracking, and to the anticipatory timing of event changes in discrete visual tracking, abundantly confirm the usefulness of the pretraining approach for the study of anticipation and encourages an interpretation in terms of mediated responses. Experiment III revealed that feedback is a necessary condition for acquisition of beneficial anticipation in our discrete tracking task, but it is not clear why deliberate feedback action was unnecessary for the transfer of anticipation in continuous tracking. Apparently the cues during pretraining for continuous tracking were distinctive enough for S to know the adequacy of his response timing, but not so in the case of the 1-sec. discrete events. And, providing feedback was present, the transfer to discrete visual tracking was independent of the visual or auditory stimulus mode used in pretraining. An acquired timekeeping capability is not wedded to the original sensory mode.

No differences were found between verbal and motor methods of pretraining for continuous tracking, and this provides us with no basis for deciding between a motor and a nonmotor locus of the mediators. While further research may uncover evidence on the locus of mediators, the absence of differences between verbal and motor methods suggests a new explanation in terms of a common type of movement-produced stimuli that accompanies all responses, whatever their locus. Proprioception is one candidate for the common factor. All of our pretraining responding had proprioceptive accompaniments, including verbal responses, and these proprioceptive traces from persisting fractional movements associated with pretraining responses may be the mediational basis of anticipatory timing in motor behavior. Time-varying proprioceptive traces may be the response-produced stimuli which function as the time perception mechanism to provide distinctive cues at time t for cuing a beneficial anticipatory overt response to appear at time $t + \Delta t$. Perhaps any pretraining response can provide transfer of anticipation in tracking through the mechanism of proprioceptive stimuli associated with it, and it does not particularly matter whether the response locus is motor-peripheral or verbal-central. An alternative common factor might be cognitive learning, and this interpretation could hold that all pretraining methods allowed S to observe the time patterning of events, and these operations of association in the perceptual field, along with confirmation of the sign-significate expectation (Tolman, 1958), would seem to be a sufficient condition for acquiring anticipatory behavior. Perhaps it would have been useful to have a pretraining condition for cognitive learning where S merely watched the events, but we rejected this at the outset because such operations do not prevent the verbally facile human S from covert verbal responding, thus producing functional equivalence to our explicit verbal pretraining. We thought it better experimental control if our verbal responses were deliberately overt and observable.

We do not deny the possibility of cognitive learning as the primary basis of pretrained anticipation in our studies, but there is the scientific problem of unequivocally specifying operations which disallow the influence of verbal responses and their stimuli, proprioceptive or

otherwise. Proprioception, on the other hand, is not quite so difficult to pursue experimentally because it can be manipulated through several defining variables (e.g., Bahrick, 1957). While evidence on uncontaminated cognition will be hard to obtain directly, a body of indirect evidence in its favor might be assembled if research revealed that the time-varying qualities of proprioception do not influence the accuracy of perceptual anticipation. If there is anything to this role for proprioception, anticipation in tracking should emerge as (*a*) some function of the physical variables of the control system associated with proprioceptive feedback, and (*b*) time between responses. Goldstone, Boardman, and Lhamon (1958) offer encouragement for the proprioceptive view with their preliminary evidence showing that children make extensive use of kinesthetic cues in originally learning to perceive time intervals.

Summary

Three experiments were performed to test the hypothesis that anticipation of directional signal change in one-dimensional continuous tracking, and time of event change in discrete tracking, is some function of mediated responses and their response-produced cues. A transfer of training design was used where practice on the criterion tracking task was preceded by a pretraining session where neutral responses were made to time regularities in the task stimuli. The design also attempted to make a first delineation of the motor vs. nonmotor locus of mediating response determinants of anticipation, if evidence was found for mediation at all.

Significant amounts of transfer of training were found and mediation as a mechanism for anticipation in tracking was found tenable. No differences were found for the motor vs. nonmotor locus of the mediators. Time-varying proprioceptive traces and cognitive learning were discussed as explanations.

References

Adams, J. A. A source of decrement in psychomotor performance. *J. exp. Psychol.*, 1955, **49**, 390–394.

Adams, J. A. Human tracking behavior. *Psychol. Bull.*, 1961, **58**, 55–79.

Adams, J. A., & Xhignesse, L. V. Some determinants of two-dimensional visual tracking behavior. *J. exp. Psychol.*, 1960, **60**, 391–403.

Arnoult, M. D. Stimulus predifferentiation: Some generalizations and hypotheses. *Psychol. Bull.*, 1957, **54**, 339–350.

Bahrick, H. P. An analysis of stimulus variables influencing the proprioceptive control of movements. *Psychol. Rev.*, 1957, **64**, 324–328.

Bartlett, F. C. Anticipation in human performance. In, *Essays in psychology: Dedicated to David Katz.* Uppsala: Almquist & Wiksells, 1951. Pp. 1–17.

Gibbs, C. B. The continuous regulation of skilled response by kinaesthetic feedback. *Brit. J. Psychol.*, 1954, **45**, 24–39. (a)

Gibbs, C. B. Movement and force in sensori-motor skill. In W. F. Floyd & A. T. Welford (Eds.), *Human factors in equipment design.* London: Lewis, 1954. Pp. 103–117. (b)

Goldstone, S., Boardman, W. K., & Lhamon, W. T. Kinesthetic cues in the development of time concepts. *J. genet. Psychol.*, 1958, **93**, 185–190.

Goss, A. E. A stimulus-response analysis of the interaction of cue-producing and instrumental responses. *Psychol. Rev.*, 1955, **62**, 20–31.

Helson, H. Design of equipment and optimal human operation. *Amer. J. Psychol.*, 1949, **42**, 473–497.

Poulton, E. C. The basis of perceptual anticipation in tracking. *Brit. J. Psychol.*, 1952, **43**, 295–302. (a)

Poulton, E. C. Perceptual anticipation in tracking with two-pointer and one-pointer displays. *Brit. J. Psychol.*, 1952, **43**, 222–229. (b)

Poulton, E. C. On prediction in skilled movements. *Psychol. Bull.*, 1957, **54**, 467–478.

Snedecor, G. W. *Statistical methods.* (5th ed.) Ames: Iowa State Coll. Press, 1956.

Tolman, E. C. *Behavior and psychological man.* Berkeley: Univer. of California Press, 1958.

Vanderplas, J. M. Transfer of training and its relation to perceptual learning and recognition. *Psychol. Rev.*, 1958, **65**, 375–385.

Vince, M. A. The part played by intellectual processes in a sensori-motor performance. *Quart. J. exp. Psychol.*, 1953, **5**, 75–86.

Vince, M. A. The relation between hand movements and intellectual activity in a skilled task. *Quart. J. exp. Psychol.*, 1955, **7**, 82–90.

(Received October 28, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 91–97

SHAPE-CONSTANCY: DEPENDENCE UPON STIMULUS FAMILIARITY[1]

C. ROBERT BORRESEN AND WILLIAM H. LICHTE

University of Missouri

The purpose of the present study was to specify, for one set of conditions, the role of learning in shape-constancy. Undertaking such an experiment does not mean that the present writers minimize the importance of stimulus variables in shape-constancy, nor does it imply that learning must be a major factor. The writers are impressed by the theory of Gibson (1950, 1959) and feel that this theory itself makes the present study desirable.

The intent of the study was not to attack the whole question of how learning influences perceptual constancy; it was limited to showing how one specific learning factor operated in one situation. Learning in this instance meant familiarity and the main feature of the study was that familiarity was manipulated as an experimental variable. The primary hypothesis tested was that the degree of shape-constancy shown in the responses to a given stimulus is a function of the amount of experience *O* has had with that stimulus. Experience (or familiarity) was crudely defined as the number of instances of being-stimulated-by-and-making-a-judgmental-response-to the object. The judgmental response arbitrarily selected was that of indicating the orientation of the stimulus. The perception of orientation (tilt or slant) has always seemed, from a logical standpoint, to be crucial in producing constancy (Beck & Gibson, 1955; Stavrianos, 1945).

On the assumption that the basic hypothesis would be supported, the experiment was also designed to evaluate the role of one condition—the *variety* of experiences with a form—in the familiarizing process. Shape-constancy is essentially the phenomenon of perceiving the same "real" shape as the retinal image pattern changes because of the different orientations of the stimulus object, therefore the *number* of orientations in an experience period could be important. The second hypothesis was that the degree of shape-constancy would be a function of this variable.

The general plan of the experiment was to use initially unfamiliar ("nonsense") figures and to have observers view-and-respond-to these different numbers of times in an experience or familiarity period. Shape-constancy for each stimulus was then to be measured in a test period. To test the second hypothesis, different groups of *S*s were to see the stimuli at different numbers of orientations in the experience period.

METHOD

Apparatus.—A sketch of the apparatus, as seen by *O*, appears in Fig. 1. The apparatus was designed to present the standard stimulus (SS) at various slants directly in front of *O* (B in Fig. 1), at a distance of 1.22 m. from his forehead; the variable stimulus (VS) used for constancy judgments was 30° to the right of the SS position (C in Fig. 1), and the circular stimulus used for orientation judgments during training was 30° to the

[1] This research was supported in part by Contract AF33(600)31315, held by International Business Machines Corporation, University of Missouri subcontract.

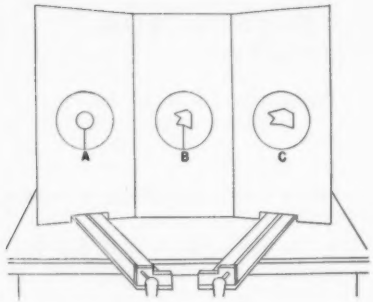

FIG. 1. Sketch of apparatus, as seen by the observer.

left of the SS position (A in Fig. 1). The *O*'s head was the center of a circle with points A, B, and C on its circumference. The *O*'s head was not restrained, although he was instructed to maintain a specified position. The *O* always used binocular vision.

For convenience, the word "stimulus" will be used in this paper to refer to the "distal" stimulus. Further, as will be seen below, the SS in this study was always a physical object (a piece of paper with a backing) while the VS was always a patch of light seen on a screen.

The *SS* shape in the middle position could be seen through a circular opening, 18 cm. in diameter, which had a removable shield between the opening and the stimulus. The shield was put in place between trials so that *O* could not see *E* changing the stimuli. A 7-w. bulb illuminated the opening so that *O* could clearly see the stimulus in the semi-dark room. The illumination on the SS was about 5 ft-c. Each SS was made of a piece of heavy photographic film faced with white paper.

On *O*'s right was the VS mechanism. The VS was produced by passing a beam of light through a piece of black photographic film with an appropriate transparent shape in the center. The beam of light then fell on a paper screen (from the rear). Light was provided by a Project-O-Chart (American Optical Company) with a 50-w. bulb, 4.5 m. from the screen. This method only approximated parallel light, but the resulting figures were not appreciably distorted or blurred. The VS mechanism was boxed in to prevent any stray light from striking the screen. The light beam could be blocked off by a shield between trials. The brightness of the patch of light constituting the VS was approximately equal to that of the illuminated white paper of the SS.

The film rectangles for producing the variable stimuli were rigidly mounted in frames which could be slipped into a holder on top of a rod in back of the paper screen. The *O* could rotate the rod and thus the film by turning a crank in front of him. By means of this rotation of the film the VS could be changed from one duplicating the shape on the film to a very narrow projection of this shape. A pointer attached to the rod supporting the film rotated over a fixed protractor so that the position of the film could be read by *E*.

On the left of *O* was a stimulus for orientation judgments obtained during the familiarity period. This consisted of a circular shape, of about the same area as the average of the standard stimuli, which could be rotated by *O* on a vertical axis through its center. A 7-w. light source illuminated this shape. Again, the illumination on the white plate was about 5 ft-c.

Stimuli.—Since the purpose of the experiment was to manipulate familiarity (number of past stimulus-response sequences), it was not possible to use the geometrical shapes usually employed in shape-constancy studies because of the previous experience of *O*s with such shapes. New shapes were therefore constructed for this experiment. They were

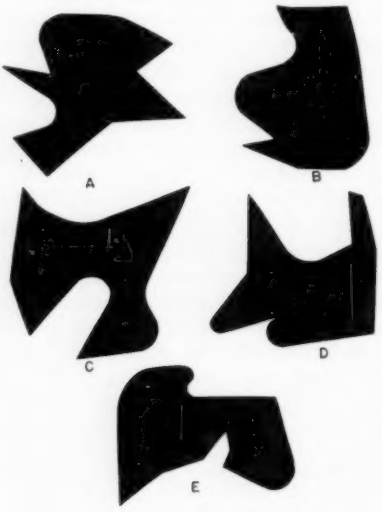

FIG. 2. The standard stimuli.

created by using a very slight modification of one of the methods (No. 4) suggested by Attneave and Arnoult (1956). The shapes employed are presented in Fig. 2.

As described previously, the VS was produced by passing a beam of light through a film. On this film the shape portion was transparent, while the surround was not. This allowed for a white projection on a black (or very dark) field. In all cases the shapes on the film were the SS shapes expanded horizontally, but not vertically. The difference was such that when the film was placed at an angle of approximately 60° from the path of the light beam the VS (on the screen) matched perfectly the shape of the corresponding SS (in the frontal plane). This relationship (film to image) was used for two reasons: (*a*) it allowed for the measurement of overconstancy, if it occurred (the VS could be wider than the physical object which was the SS); and (*b*) it avoided a suggestive perceptual landmark on the VS continuum. If the shape on the film had been the same as the SS, the VS on the screen would, as *O* turned the crank, expand to the point of equality with the objective shape of the *SS* and then begin to contract. This point of change would represent perfect shape-constancy.

Subjects.—The *O*s were all from undergraduate psychology classes at the University of Missouri. There were no restrictions made with regard to wearing glasses, age, or sex. There were 40 *O*s in all, 15 in each of the two experimental groups and 10 in a control group.

Procedure.—There were two parts to the experiment, a familiarity period and a constancy period. During the first, *O*s made judgments about the orientation of the SS; during the second they made shape-constancy judgments.

There were two experimental variables in the familiarity part of the experiment: (*a*) frequency of presentation, and (*b*) the orientation of the shapes.

Each *O* had 40 familiarity trials in which the five stimuli were presented at different frequencies. For any given *O* the five stimuli were presented at frequencies (one stimulus at each frequency) of 0, 4, 8, 12, and 16. There were five separate combinations of frequency-of-presentation and shapes; in other words, all shapes were rotated through all frequencies of presentation in a latin square design.

The second variable in the training period was the orientation of the *SS*. (The orientation or position of the stimulus was measured with respect to the line of sight from *O* to the stimulus, 0°. When the stimulus was in the frontal-parallel plane the edge to the left of *O* was at 90°.) In terms of the orientation variable there were two groups of *O*s. In Group I *O*s had one-half the presentations of each shape at 110° and one-half at 140°. Group II had one-fourth of the presentations of each stimulus at 110°, one-fourth at 120°, one-fourth at 130°, and one-fourth at 140°. This design allowed the absolute frequency of any particular stimulus to be the same under all conditions. For example, if Stimulus B were to be presented eight times, it would appear in two orientations (110° and 140°) four times each in Group I, but in four orientations (110°, 120°, 130°, and 140°) only two times each in Group II.

A control group of 10 *O*s (Group III) did not have any familiarity trials.

On each familiarity trial *O* was to indicate the orientation of the SS by rotating the circle on his left until he thought that it matched perfectly the orientation of the SS. This judgment was independent of shape, as the standard stimuli were the nonsense shapes whereas the orientation shape was a circle. The order of presentation of the standard stimuli was randomly determined for each condition and that order, and its inversion, were used in a counterbalanced manner.

The essentials of the instructions given *O* at this point were that the experiment involved judgments about abstract shapes, that various shapes would be placed in the middle window at some angle of slant, and that the task was to adjust the slant of the circle in the left window until it duplicated that of the object in the center window. The *O* was to face towards a spot halfway between the two windows and keep the head in that position.

After the familiarity period *O* had a few minutes rest, and then went on to the shape-constancy part of the experiment. Each *O* (including the control *O*s) had 30 constancy trials, 6 for each of the five stimuli. Each SS was placed at an orientation of 135° for three trials and at 155° for three trials. The use of different orientations was primarily to prevent *O*s from becoming too fixed in their responses by recognizing that the SS always appeared in the same orientation. These particular values were chosen on the assumption that they would normally (with familiar figures) give Brunswik ratios in the upper-middle range (Lichte, 1952).

It was *O*'s task during the constancy trials

to view the rotated SS and adjust the VS film plate until he thought that the VS image on the screen was the same shape as the SS when at 90°, i.e., the objective shape of the SS. The *O*s rotated the VS film, thus changing the VS image, by turning a crank. For one-half of the trials of each shape the VS was set to give an image narrower than the frontal plane projection of the SS, and one-half the trials, wider. After *O* had made an adjustment to his satisfaction his score was read from a pointer attached to the VS rod. During the constancy trials, two different random orders of presentation of the SS were used in a counterbalanced manner.

The substance of the instructions given *O* for the second part of the experiment was that he would see various shapes (objects) in the middle window, and in the right window he would see patches of light on the translucent screen. The patch of light could be changed in shape by turning a crank and *O*'s task was to adjust its shape to duplicate that of the plate in the middle window, when the latter was considered as an object independent of its slant; *O* was to face midway between the two windows.

Results and Discussion

Transformation of scores.—As the film was manipulated the shape of the VS on the screen changed continuously in width, but remained the same in height. Each shape on the continuum was associated with one image width. The measurement of image width was therefore considered a satisfactory index of the physical variable of shape. The original data for *O*s' responses were in terms of the angle of the film plate with reference to a line perpendicular to the screen. By means of the appropriate trigonometric functions these angles were transformed into width measures for the VS on the screen.

There were three trials for each SS presentation and therefore three width scores. The means of these widths were used to compute a BR (Brunswik ratio). There were 10 BRs for each *O*. These data were treated by an analysis of variance. The particular test used was a combination of a factorial and latin square design.

TABLE 1

Analysis of Variance of Brunswik Ratios

Source	*df*	*MS*	*F*
Between *Ss*	29	.23	
Orientation	1	.01	
$S \times F_{(b)}$	4	.42	1.91
$S \times F \times O_{(b)}$	4	.15	
$Error_{(b)}$	20	.22	
Within *Ss*	120	.11	
Shapes	4	1.02	20.40**
Frequency	4	.43	8.60**
$S \times F_{(w)}$	12	.11	2.21*
$S \times O$	4	.02	
$F \times O$	4	.11	2.20*
$S \times F \times O_{(w)}$	12	.18	3.60**
$Error_{(w)}$	80	.05	

* $P < .05$.
** $P < .01$.

The complete analysis of variance is presented in Table 1. Frequency of presentation as a main effect (*F*) was significant at better than the .01 level. The mean BRs for frequencies of 0, 4, 8, 12, and 16 were .46, .66, .61, .77, and .74, respectively. These results in general support the main hypothesis that shape-constancy is an increasing function of shape familiarity where familiarity is defined as the number of times that a shape was seen and responded to by *O*. This function, separated out for Groups I and II, is presented in Fig. 3.

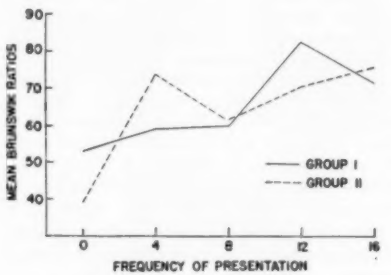

Fig. 3. Brunswik ratios for each frequency, by group.

The writers ascribe considerable theoretical importance to this basic finding. The function obtained is considered important, but so also is the obtained BR of .46 for the zero frequency-of-presentation condition. It seems reasonable to conclude from this that the stimulus factors discussed by Gibson (1950, 1959) are operating to produce this much shape-constancy.

A comparison of the mean BR of .43 for the control group of 10 *O*s (with no familiarity trials) with the zero frequency-of-presentation condition for the experimental groups indicated that there was no significant difference between the two conditions. This result is interesting in that it may be interpreted as evidence against any generalization or positive transfer in the experimental groups from the four stimuli in the familiarity period to the fifth stimulus, seen first in the second period. If there had been positive transfer one would expect the zero frequency-of-presentation condition of the experimental groups to have a somewhat higher BR than the control group.

An analysis was run comparing the 4, 8, 12, and 16 levels of frequency of presentation. The results of this test indicated no significant difference between these levels of presentation. The *t* test was applied to all possible pairs of conditions and only the 0–12 and 0–16 differences were significant at the .01 level. This may mean that the zero frequency-of-presentation factor is contributing the major portion of the variance which resulted in frequency of presentation being a significant variable. This function deserves further study, however, for in a nearly-identical previous study (Borresen & Lichte, 1959) with 30 *O*s a significant function was found when the frequencies of presentation were 2, 4, 8, 12, and 16.

The number of orientations (0) of the SS during the familiarity session was not significant; the mean BR for Group I was .65 and for Group II, .64. A glance at Fig. 3 clearly indicates the extensive overlap between these two groups. Thus, the secondary hypothesis, that shape-constancy increases with the number of different orientations, was not supported. However, the conditions of the experiment gave a very limited approach to the subject of orientation and the relations here seem worthy of further study.

Shape as a main effect (S) was significant at better than the .01 level. For Stimuli A, B, C, D, and E the BRs were .63, .50, .51, .65, and .95, respectively. A subsidiary analysis comparing shapes A, B, C, and D indicated no significant difference between these shapes. It is of additional interest to note that in the control group also, Stimulus E had a very high BR. In this group the mean BRs for Stimuli A, B, C, D, and E were .28, .33, .29, .42, and .80, respectively.

This points up a rather interesting situation. The shapes used in this study were all constructed by the same set of rules; they are, in a sense, all from the same physical family. Yet the data indicate that psychologically they are not the same. Other studies, with dependent variables such as shape recognition or discrimination (Arnoult, 1954; Fitts, Meyer, Rappaport, Anderson, & Leonard, 1956), have pointed toward the same conclusion: that shapes of a given physical family are not necessarily of the same psychological family.

The importance of the shape characteristic of the stimuli and the

TABLE 2

MEAN BRUNSWIK RATIO FOR EACH SHAPE AT EACH FREQUENCY

Shapes	Frequencies				
	0	4	8	12	16
A	.38	.56	.60	1.02	.60
B	.41	.47	.43	.36	.82
C	.55	.44	.43	.52	.62
D	.19	.86	.58	.79	.85
E	.77	1.00	1.01	1.17	.81

complexity of the whole subject of the relation of constancy to familiarity is indicated by the fact that the Shape × Frequency interaction is significant at less than the .05 level. The means relating to this interaction are shown in Table 2. A great deal of chance variability is undoubtedly represented here, perhaps particularly in the case of Stimulus D, but the table suggests the importance of further study of this interaction.

The perceptual constancies are undoubtedly most complete in realistic or everyday situations. In this study the stimulus situation was artificial and rather impoverished and thus might be expected to produce incomplete constancy. In addition, the nature of the instructions to *O* was not such as to produce the greatest degree of constancy. The latter conclusion comes from a subsequent study on the effect of instructions (Lichte & Borresen, 1960). As a result, at least in part, of these conditions incomplete constancy was found and this may have increased the probability of finding, or the size of, the hypothesized relationship between familiarity and degree of constancy.

In other studies with somewhat different conditions (Sheehan, 1938; Stavrianos, 1945) high degrees of constancy or overconstancy have been obtained. These prove little with regard to the point at issue, however, because the stimuli were rectangles, ellipses, and diamonds, and were therefore highly familiar at the outset of the experiment.

SUMMARY

The present experiment tested the hypothesis that shape-constancy is a positive function of shape familiarity, where familiarity is defined as the number of times *O* is allowed to view-and-respond-to a shape during a familiarity period. Some effects of the orientation of the standard stimulus during a familiarity period were also tested.

The stimulus materials were five moderately complex shapes systematically constructed from randomly determined coordinates. During a familiarization period the five stimulus shapes were presented 0, 4, 8, 12, and 16 times, one stimulus at each frequency, and *O*s were asked to indicate the orientation of the SS by matching a circular shape to that orientation. There were two familiarization conditions, each having a different number of orientations of the SS (two for Group I, $N = 15$, and four for Group II, $N = 15$). A control group ($N = 10$) received no familiarity trials.

In the second part of the experiment constancy was measured, using the same five forms. The *O*s were asked to set the VS to match the real shape of the SS. Two orientations were used.

The results supported the main hypothesis that shape-constancy is a positive function of shape familiarity. Shape-constancy for the zero frequency-of-presentation condition for the experimental groups was not significantly different from that of the control group. Shape as a main effect was significant, with most of the variance coming from one shape, and the interaction of shape and frequency was significant. Orientation during the familiarity period was not significant. The number of orientation positions, within a given range, seems to be unimportant.

REFERENCES

ARNOULT, M. D. Shape discrimination as a function of the angular orientation of the stimuli. *J. exp. Psychol.*, 1954, **47**, 323–328.

ATTNEAVE, F., & ARNOULT, M. D. The quantitative study of shape and pattern perception. *Psychol. Bull.*, 1956, **53**, 452–471.

BECK, J., & GIBSON, J. J. The relation of apparent shape to apparent slant in the perception of objects. *J. exp. Psychol.*, 1955, **50**, 125–133.

BORRESEN, C. R., & LICHTE, W. H. Shape constancy as a function of form familiarity.

Paper read at Midwestern Psychological Association, Chicago, May 1959.

FITTS, P. M., MEYER, W., RAPPAPORT, M., ANDERSON, N., & LEONARD, J. A. Stimulus correlates of visual pattern recognition: A probability approach. *J. exp. Psychol.*, 1956, **51**, 1–11.

GIBSON, J. J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950,

GIBSON, J. J. Perception as a function of stimulation. In S. Koch (Ed.), *Psychology: A study of science.* Vol. 1. New York: McGraw-Hill, 1959. Pp. 456–501.

LICHTE, W. H. Shape constancy: Dependence upon angle of rotation; individual differences. *J. exp. Psychol.*, 1952, **43**, 49–57.

LICHTE, W. H., & BORRESEN, C. R. The influence of instructions on degree of shape constancy. Paper read at Midwestern Psychological Association, St. Louis, May 1960.

SHEEHAN, M. R. A study of individual consistency in phenomenal constancy. *Arch. Psychol.*, Geneve, 1938, **31**, No. 222.

STAVRIANOS, B. K. Relation of shape perception to explicit judgments of inclination. *Arch. Psychol.*, Geneve, 1945, **42**, 1–94.

(Received October 31, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 1, 98–104

LATENT EXTINCTION AS A FUNCTION OF NUMBER AND DURATION OF PRE-EXTINCTION EXPOSURES [1]

JAMES A. DYAL[2]

Southern Methodist University

The latent extinction effect is a decreased resistance to extinction resulting from direct placement of *S*s into nonreinforced contact with previously positive stimuli prior to regular extinction trials. The weight of the evidence provides convincing support for the reliability of the effect since 10 out of 12 experiments have yielded positive results. In addition to demonstrating the effect, a few of these studies have attempted to determine the influence of other variables on latent extinction. Moltz and Maddi (1956) investigated the drive variable at three levels of hunger (0, 22, and 44 hr.). Their data suggested a negatively accelerated function relating drive level and number of correct choices during extinction. Pliskoff (1955) obtained similar results using 0, 8, and 22 hr. of hunger. He also investigated number of latent extinction exposures (1, 4, and 20) and found that this variable influenced choice behavior only if the placements were given in the same goal box as used during training. A parametric study by Thomas (1958) explored the effects of duration of a single goalbox placement (1, 5, or 15 min.) and interval between goalbox placements and extinction (15 sec., 5 min., or 15 min.). The purpose of the present experiment was to investigate the effects of duration and number of goalbox placements on the latent extinction effect.

METHOD

Subjects and apparatus.—The *S*s were 96 naive Sprague-Dawley male rats between 80 and 100 days old at the start of training. The apparatus was a Y maze, 5¾ in. deep and covered with ¼-in. hardware cloth. Upon leaving the 11 × 4 in. neutral gray startbox, *S* encountered a choice area shaped as a truncated triangle and having the following dimensions: base, 24 in., top, 10½ in., and sides, 13¾ in. The runway arms were both 5½ × 4½ in. A left turn led to a flat white runway having 36 nickel-plated, round-head thumbtacks set haphazardly in the floor. The circular goalbox was also painted flat white and was 10½ in. in diameter. Its sides were constructed of stainless steel and the wooden floor was covered with a removable piece of ¼-in. hardware cloth. A right turn led through an unpainted runway to an unpainted triangular goalbox having 13-in. sides. The *S*s thus had available brightness, tactual, and extramaze cues to facilitate discrimination of the arms and goalboxes. The floor of each goalbox contained a ¼-in. deep food well which prevented *S* from seeing the food until he entered the goalbox. The neutral box into which the control *S*s were placed was 10 in. square and painted flat black. The apparatus was instrumented with a photoelectric system which permitted the recording of starting latencies, choice times, running times, and goalbox times on .001-min. Standard Electric timers. The goalbox times (GT) indicate the time required to traverse the last 2½ in. of the runway and the first 8½ in. of the goalbox.

Pretraining.—The *S*s were adapted to a 23-hr. food deprivation schedule for 7 days prior to the start of training. On Pretraining Day 5, *S*s were permitted to explore the apparatus in groups of 6 for 10 min. On Day

[1] The author wishes to thank Richard Carpenter and Bruce Patterson who ran the animals for this experiment. This research was supported by Public Health Service Research Grant M-3331 from the National Institute of Mental Health.

[2] Now at Texas Christian University.

6, each *S* was given 60 sec. contact eating following direct placement into each goalbox. The *S* was then given five free-choice trials with no food to determine turning preferences. On Day 7, each *S* received three massed forced choices to food on the nonpreferred side. Eating was allowed for 60 contact sec. on the forced trials.

Training.—The *S*s were trained in groups of 6 and were given five free-choice trials per day. The food (one brick of Purina laboratory chow) was always on the nonpreferred side. Following a correct choice, *S* was permitted to eat in the goalbox during the period when *E* was recording the time data. This period averaged 10–15 sec. The intertrial interval was never less than 5 min., which time *S*s spent in wire cages with water continuously available.

A multiple criterion of learning was established: (*a*) each *S* must have had a minimum of 25 training trials; (*b*) 9 of the last 10 trials must have been correct; and (*c*) the error if any, must not have been on the first trial of the last day of training.

Prior to reaching criterion, each *S* was randomly assigned to one of 12 groups. Nine experimental *S*s in each replication were assigned to one of the combinations of duration and number of placements represented in Fig. 1. Three *S*s were assigned to control groups having the duration and number values of the diagonal cells. The experiment was run in eight replications by 2 different *E*s. For convenience of exposition, the specific combination of duration, number, and experimental vs. control will be designated by group numbers, e.g., Group 6 refers to *S*s receiving 10 placements of 60 sec. duration in the goalbox, whereas 5C refers to 5 placements of 60 sec. duration in the neutral box. The effects of the independent variables of number and duration are present in Cells 1 through 9 and this will be referred to as the 3 × 3 design. The presence of the latent extinction effect is represented in the experimental and control portions of Cells 1, 5, and 9; the 2 × 3 design.

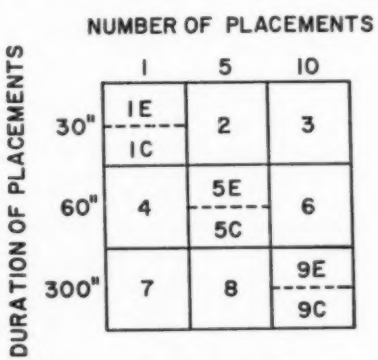

FIG. 1. Experimental design.

Latent extinction and regular extinction.—Twenty-two to 24 hr. after reaching the criterion each experimental *S* was given the appropriate number and duration of placements into the formerly baited goalbox. The interplacement interval of 60 sec. was spent in the retraining cages and in transit. Sixty to 90 sec. following the last goalbox placement, each *S* was given the first of 10 regular extinction test trials. The intertrial interval ranged from 5 min. early in extinction to 10 min. late in extinction. The procedure was the same for the control *S*s except they were placed into a neutral box rather than the goalbox.

RESULTS

In order to interpret the parametric data from the experimental *S*s in the 3 × 3 design as being relevant to latent extinction, it is first necessary to demonstrate that latent extinction actually occurred. The relevant data are provided by comparison of the experimental and control groups in the 2 × 3 design.

2 × 3 Design

Goalbox times.—On the basis of a pilot study it was determined that the dependent variable which was likely to be the most sensitive to the effects of goalbox placements was the goalbox time (GT) on the first correct trial.[3] Because of apparent heterogeneity of variance of the raw data, each GT score was transformed by log (1 + GT). Bartlett's test pro-

[3] The running times, choice times, and starting times indicated the same general trends as the goalbox times but the E-C differences were small and became smaller the further the segment was from the goal area. The data from these measures were not analyzed further.

TABLE 1

SIGNIFICANT RESULTS OF ANALYSIS OF VARIANCE FOR LOG (1 + GT) AND ERROR SCORES IN VARIOUS SEGMENTS OF EXTINCTION (2 × 3 DESIGN)

Extinction Phase	Source	*df*	*MS*	*F*
	Log (1 + GT) Scores			
1st correct trial	Exp. vs. Control	1	.3601	23.38****
Correct choices Trials 1–5	Exp. vs. Control	1	.1202	13.07****
	Exposure time	2	.2056	21.64****
	Exp. vs. Control × Exposure time	2	.0580	6.10****
Correct choices Trials 1–10	Exp. vs. Control	1	.0453	6.66***
	Exposure time	2	.0378	5.55***
	Error Scores			
Trials 1–5	Exp. vs. Control	1	.1480	7.18**
Trials 1–5	Exp. vs. Control × Exposure time	2	.0825	4.00*
Trials 1–10	Exp. vs. Control	1	.1660	4.61*
Trials 1–10	Exp. vs. Control × Exposure time	2	.1460	4.06**

* $P < .05$.
** $P < .025$.
*** $P < .005$.
**** $P < .001$.

vided no basis for the rejection of the hypothesis of homogeneity of variance with regard to the transformed data.

Results of analyses of variance applied to each of the extinction segments as well as to all 10 trials combined, are presented in the upper part of Table 1. Only those variables which contributed significant variance are included; also note that GTs are based on correct trials only. It may be seen that the experimental-control differences are highly significant during all extinction segments except Trials 6–10. Subgroup comparisons based on *t* tests in which the error term was derived from the mean square for error from the analysis of variance support the conclusion that the latent extinction effect is primarily related to 5E-5C and 9E-9C differences. In all extinction phases Groups 5E and 9E took significantly longer to enter the positive goalbox than the comparable control groups. The differences for Groups 1E and 1C, although in the appropriate direction, were not statistically reliable. The size of the latent extinction effect, although still significant when all 10 extinction trials are used as the unit of analysis, tends to diminish with continued extinction; there were no reliable differences between any of the experimental and control groups on Extinction Trials 6–10. This view is further supported by Fig. 2 in which GTs on correct trials are combined for all experimental and control groups. It should be noted that the experimental *S*s show an unusual *decrease* in time required for goalbox entry as extinction continues, whereas the control *S*s manifest the usual increasing latency.

With reference to Table 1 the experimental condition designated as exposure time represents the total amount of time spent in the goalbox during latent extinction placements. For Group 1 this value was 30 sec. (1 placement of 30 sec. duration); Group 5 had a total exposure of 300

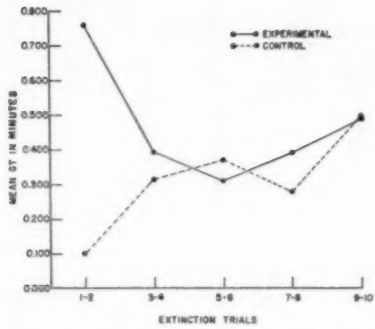

FIG. 2. Mean goal times for all experimental and all control groups combined, in blocks of two extinction trials.

sec. (5 × 60 sec.), and Group 9 had an exposure of 3000 sec. (10 × 300 sec.). Subgroup comparisons indicated that Group 1E entered the goalbox significantly faster than either Group 5E or Group 9E during the first 5 trials of extinction and over all 10 extinction trials. In no case was Group 5E significantly different from Group 9E.

Errors.—An error is defined as a choice of the previously unbaited goalbox. Evaluation of the number of errors made by the subgroups on Test Trial 1 by a test for uncorrelated proportions indicated that only the 1E-1C difference was reliable ($z = 2.00$; $P < .025$) although the 9E-9C difference approached significance ($z = 1.51$; $P < .07$).

Error scores for Trials 1–5, 6–10, and 1–10 were based on transformation of the number of errors by $\sqrt{x} + \sqrt{1 + x}$ (Edwards, 1960, p. 128) and were analyzed by analysis of variance. The significant F ratios resulting from each of these analyses are presented in the lower part of Table 1. The experimental vs. control variable is significant over all trials except 6–10. The significant interaction was due to the reversal in number of errors for Groups 1E and 1C.

Another index of performance based on error scores is the trial on which the first error was made. Analysis of variance applied to these scores transformed by $\sqrt{x} + \sqrt{1 + x}$ resulted in a highly significant difference between experimental and control conditions ($F = 22.73$, $df = 1/42$, $P < .001$).

3 × 3 Design

Goalbox times.—The mean log (1 + GT) as a function of number of placements and duration of each placement is represented in Fig. 3. A summary of the significant effects revealed by analysis of variance applied to these data for the various segments of training is presented in Table 2. Both duration and number of goalbox placements influence GT in all segments of extinction except Trials 6–10. Duration has its greatest effect on Trial 1 and does not influence Trials 1–5 except in interaction with number, whereas the main effect attributable to number of placements is present throughout the first 5 trials. It may be seen that the strength of the effect during Extinction Trials 1–5 is sufficiently great that when all 10 trials are combined both variables significantly influence GT.

TABLE 2

SIGNIFICANT RESULTS OF ANALYSIS OF VARIANCE OF LOG (1 + GT) FOR VARIOUS PHASES OF EXTINCTION (3 × 3 DESIGN)

Extinction Phase	Source	*df*	*MS*	*F*
1st correct trial	Duration (D)	2	.1684	5.36***
1st correct trial	Number (N)	2	.1704	5.34***
Correct choices	Number (N)	2	.1553	25.88****
Trials 1–5	N × D	4	.1544	25.73****
Correct choices	Duration	2	.0910	12.13***
Trials 1–10	Number	2	.0759	10.17***

*** $P < .005$.
**** $P < .001$.

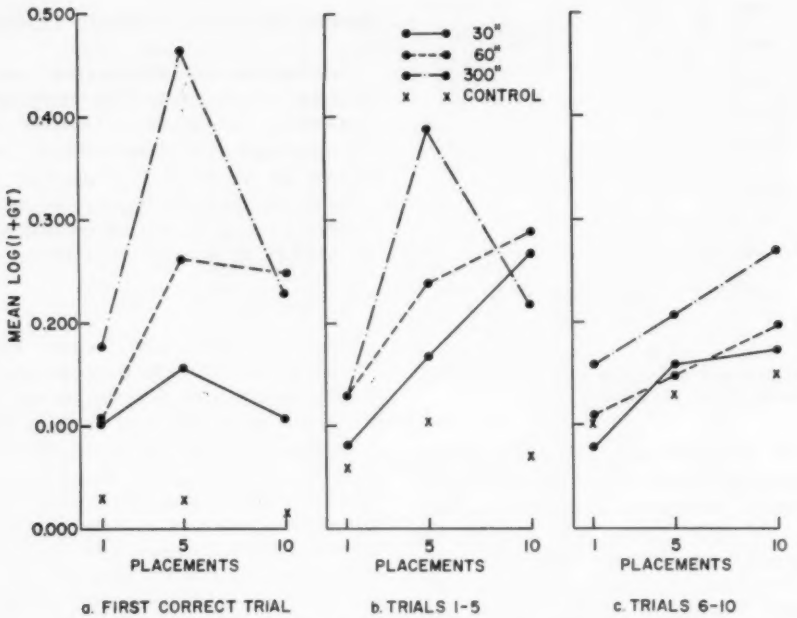

FIG. 3. Mean log (1 + GT) during three extinction phases as a function of duration of placement.

Appropriate subgroup comparisons of the data represented in Fig. 3 and Table 3 reveals that Group 8 (5×300 sec.) was significantly ($P < .005$) slower in entering the goalbox on the first correct test trial than was Group 7 (1 × 300 sec.) or Group 9 (10 × 300 sec.). Group 8 was also significantly ($P < .001$) slower than Group 5 (5 × 60 sec.). During Trials 1–5 those groups which were given 30-sec. placements differed reliably as a function of the number of placements; Group 1 < Group 2 < Group 3 ($P < .025$, $P < .005$). Group 5 (5×60 sec.) was significantly ($P < .005$) slower than Group 4 (1 × 60 sec.). For those groups having a duration parameter of 300 sec., Group 8 was reliably slower than Group 7 ($P < .001$) and Group 9 ($P < .001$). During Trials 1–10, Group 6 (10 × 60 sec.) was slower than Group 3 (10 × 30 sec.) and Group 8 was slower than Group 5 (both Ps $< .005$). Two significant differences were obtained for the number variable; Group 2 (5 × 30 sec.) was slower than Group 1 (1 × 30 sec.) and Group 8 was slower than Group 7 ($P < .05$, $P < .001$).

Errors.—Analyses of variance and t tests were applied to the transformed error scores for various phases of extinction. With the exception of a significant interaction with duration for the last half of extinction, it is only the number variable which significantly influenced choice behavior. Subgroup comparisons indicated that the effect was dependent primarily on the significantly greater number of errors made by Group 8 than by Groups 7 or 9.

Discussion

The data indicate the presence of a strong latent extinction effect, thus further supporting the reliability of the phenomenon. The magnitude of the effect is determined by both number and duration of placements, sometimes independently and sometimes in interaction depending on the response measure and the phase of extinction. Considering both dependent variables, the independent variable of number of placements has a greater effect on extinction than does duration of the placement. However, the effects of both independent variables are transitory, being greatest during the first extinction trial.

The present results are somewhat opposed to those reported by Thomas (1958) with regard to the duration variable. He found that duration affected the error scores but not running times; whereas, in the present experiment GT was influenced by duration but error scores were not. If we assume that stimuli in closest association with the consummatory response will be most affected by nonreinforced placements then Thomas' choice measure is likely to have been more sensitive since his choice area was considerably closer to the goalbox than was the case in the present study (12 in. vs. 48½ in.). Similarly, however, the present speed measures (GT) should be more sensitive to the effects of goalbox placements than Thomas' running speed score. This view is supported by another experiment in the present series in which it was found that when measures of running speed were available for five segments of a straight runway it was *only* performance in the immediate goalbox area which was influenced by latent extinction procedures.

In the only other study dealing with the effects of number of placements, Pliskoff (1955) found that error scores were significantly influenced by number of placements throughout 20 extinction trials; whereas, in the present data the major effect of number was on the first 5 trials. He gave his *S*s 1, 4, or 20 exposures of 30-sec. duration. His groups are thus most comparable to Groups 1, 2, and 3 in which we also find an increasing trend relating errors to number of placements.

The most complete theoretical account of latent extinction per se has been provided by Moltz (1957) and is based on the diminution of the fractional anticipatory goal response resulting from nonreinforced elicitation of r_g during pre-extinction placements. The r_g framework would predict that the greater the number and duration of exposures in the empty goalbox the greater the reduction of r_g and the more *S*s would slow down in their approach to the goalbox. Also, the negatively accelerated relation between GT on the first correct trial and total exposure time would be expected by r_g theory. However, it is difficult to see how the concept of increasing r_g diminution could be invoked to account for the obtained *decrease* in GT exhibited by the experimental groups over the 10 extinction trials. On the contrary, r_g theory would predict a continued increase in running time as regular extinction proceeds.

The fact that the difference between experimental and control *S*s disappears by the end of 3–4 trials is quite consistent with a frustration induction theory of latent extinction. If we assume that nonreward in the presence of previously rewarded stimuli results in an emotional state of frustration, then this aversive drive should still be present 60 sec. after the last placement (Trial 1) and this would result in *S*s being reluctant to enter the goal area. However, such an emotional response would be expected to dissipate fairly quickly, perhaps in a matter of 15 min. (Thomas, 1958) after being returned to the home cage, and the resulting bow-shaped curve would be expected for the experimental groups. A more detailed examination of the performance changes from Trials 1–2 to Trials 3–4 indicates that the greatest shift occurred for those groups which had the greatest total exposure. This is also consistent with a frustration induction view in that the longer the

exposure the greater the amount of frustration on Trial 1.

An implication of the present experiment for future studies of latent extinction is that the condition of five placements of 60-sec. duration (Group 5) is probably the condition of choice. It is economical of E time, and has the further advantage of being the approximate value on which the bulk of the previous literature is based.

SUMMARY

Ninety-six rats were trained to criterion in a Y maze and then given differential treatment with regard to number (1, 5, 10) and duration (30, 60, 300 sec.) of unrewarded goalbox placements. Control *Ss* were run in the diagonal cells of the 3×3 design and differed from the experimental *Ss* only in that they were placed into a neutral box.

A strong latent extinction effect was obtained for the dependent variables of speed of entering the goalbox (GT) and number of errors. The effect was strongest on the first correct trial in extinction and became progressively weaker as extinction continued. The magnitude of the effect was influenced by number of placements, duration of the individual placements, and the total exposure time, sometimes independently and sometimes in interaction depending on the response measure and the phase of extinction.

The results were compared with those obtained by previous investigators and two theories of latent extinction were examined. Both theories, frustration and r_g diminution, could accommodate the results pertaining to the effects of number and duration of placements. However r_g theory has difficulty explaining the transitory nature of the effect.

REFERENCES

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1960.

MOLTZ, H. Latent extinction and the fractional anticipatory response mechanism. *Psychol. Rev.*, 1957, **64**, 229–241.

MOLTZ, H., & MADDI, S. R. Reduction of secondary reward value as a function of drive strength during latent extinction. *J. exp. Psychol.*, 1956, **52**, 71–76.

PLISKOFF, S. S. Response elimination as a function of generalization, motivation, and number of nonresponse extinction trials. Unpublished doctoral dissertation, New York University, 1955.

THOMAS, A. R. Some variables affecting latent extinction. *J. exp. Psychol.*, 1958, **56**, 203–212.

(Received October 31, 1960)

ERRATUM

In the article "Role of Instructions in Two-Choice Verbal Conditioning with Contingent Partial Reinforcement" by J. Koehler, Jr., (*J. exp. Psychol.*, 1961, **62**, 122–125), the line following the asymptotic expressions (p. 122) should read as follows: "It can be demonstrated algebraically that the first asymptotic expression is larger than the second one when $\pi_1 > \frac{\pi_2(1 - \pi_2)}{1 - \pi_1}$ and $\pi_1 < 1$."